GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 6, 2005, 16:01:23 ; Search time 735.656 Seconds
(without alignments)
1383.200 Million cell updates/sec

Title: US-10-729-421-40
Perfect score: 21
Sequence: 1 cagtgacatgcaggtctagct 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		19.4	92.4	232064	2	AC109544	AC109544 Rattus no
2	c	19.4	92.4	258823	2	AC133226	AC133226 Rattus no
3	c	18.4	87.6	140394	2	CR352267	CR352267 Danio rer
4	c	18.4	87.6	155858	2	AP000772	AP000772 Homo sapi
5	c	18.4	87.6	161516	2	CR391906	CR391906 Danio rer
6		17.8	84.8	40392	3	U21308	U21308 Caenorhabdi
7	c	17.8	84.8	72655	2	AC100344	AC100344 Mus muscu
8	c	17.8	84.8	116076	10	AL831718	AL831718 Mouse DNA
9	c	17.8	84.8	125354	2	AC148761	AC148761 Medicago
10		17.8	84.8	139884	10	AC110379	AC110379 Mus muscu
11	c	17.8	84.8	144142	10	AC102916	AC102916 Mus muscu
12	c	17.8	84.8	162289	4	AC097230	AC097230 Sus scrof
13	c	17.8	84.8	162560	2	AC069005	AC069005 Homo sapi
14	c	17.8	84.8	167878	9	AC103719	AC103719 Homo sapi
15	c	17.8	84.8	189662	9	AC015468	AC015468 Homo sapi
16		17.8	84.8	197796	2	AC129792	AC129792 Rattus no
17	c	17.8	84.8	197796	2	AC129792	AC129792 Rattus no
18		17.8	84.8	203690	2	AC087221	AC087221 Homo sapi
19		17.8	84.8	214765	10	AC115746	AC115746 Mus muscu
20	c	17.8	84.8	222540	2	AC120123	AC120123 Mus muscu
21		17.8	84.8	260600	2	AC115307	AC115307 Rattus no
22	c	17.8	84.8	281447	2	AC129380	AC129380 Rattus no
23	c	17.4	82.9	5446	6	BD185177	BD185177 Novel gen
24	c	17.4	82.9	46070	2	AC121561	AC121561 Homo sapi
25		17.4	82.9	85566	9	AL133227	AL133227 Human DNA
26		17.4	82.9	89171	9	AC008404	AC008404 Homo sapi
27	c	17.4	82.9	92946	2	AC140020_3	Continuation (4 of
28	c	17.4	82.9	93714	2	AL161661	AL161661 Homo sapi
29	c	17.4	82.9	110000	2	AC139803_1	Continuation (2 of
30		17.4	82.9	118499	9	AC121562	AC121562 Homo sapi
31		17.4	82.9	134878	9	AC140847	AC140847 Homo sapi
32		17.4	82.9	146174	2	AC138823	AC138823 Homo sapi
33		17.4	82.9	146597	2	AC121323	AC121323 Homo sapi
34		17.4	82.9	154957	2	AC080126	AC080126 Homo sapi
35	c	17.4	82.9	159347	2	AC138971	AC138971 Homo sapi
36		17.4	82.9	170261	2	AC008542	AC008542 Homo sapi
37		17.4	82.9	171398	2	AC141597	AC141597 Homo sapi
38	c	17.4	82.9	172276	9	AC139795	AC139795 Homo sapi
39	c	17.4	82.9	174164	9	AC116166	AC116166 Homo sapi
40		17.4	82.9	175612	2	AC141254	AC141254 Homo sapi
41	c	17.4	82.9	176276	2	AC141255	AC141255 Homo sapi
42		17.4	82.9	181697	9	AC116165	AC116165 Homo sapi
43		17.4	82.9	182715	9	AC138819	AC138819 Homo sapi
44	c	17.4	82.9	185796	2	AC138857	AC138857 Homo sapi
45		17.4	82.9	185986	2	AC140123	AC140123 Homo sapi
46	c	17.4	82.9	188977	2	AC140146	AC140146 Homo sapi
47	c	17.4	82.9	189076	2	AC023788	AC023788 Homo sapi
48	c	17.4	82.9	190874	2	AC139830	AC139830 Homo sapi
49		17.4	82.9	191802	2	AC136614	AC136614 Homo sapi
50	c	17.4	82.9	192993	2	AC145108	AC145108 Homo sapi
51	c	17.4	82.9	196507	2	AC135177	AC135177 Homo sapi
52	c	17.4	82.9	197866	2	AC138895	AC138895 Homo sapi
53	c	17.4	82.9	197972	9	AC138865	AC138865 Homo sapi
54	c	17.4	82.9	201638	2	AC144987	AC144987 Homo sapi
55	c	17.4	82.9	216546	2	AC138863	AC138863 Homo sapi
56	c	17.4	82.9	222726	9	AC135069	AC135069 Homo sapi
57	c	17.4	82.9	236723	2	AC112477	AC112477 Rattus no
58	c	17.4	82.9	314133	2	AC129914	AC129914 Homo sapi
59	c	17	81.0	2862	10	AF348447	AF348447 Mus muscu
60	c	17	81.0	124034	9	HS416J7	AL035696 Human DNA
61		17	81.0	141596	10	AL603923	AL603923 Mouse DNA
62	c	17	81.0	182109	2	AC140817	AC140817 Homo sapi
63		17	81.0	188920	2	AC140709	AC140709 Homo sapi
64		17	81.0	189805	9	AC004169	AC004169 Homo sapi
65		17	81.0	192355	10	AC124126	AC124126 Mus muscu
66	c	17	81.0	196835	2	AC142538	AC142538 Homo sapi
67		17	81.0	200000	2	AC004630	AC004630 Homo sapi
68		17	81.0	206252	2	AC141064	AC141064 Homo sapi
69	c	17	81.0	206833	10	AC124178	AC124178 Mus muscu
70		17	81.0	207436	2	AC140823	AC140823 Homo sapi
71		16.8	80.0	1265	10	BC016537	BC016537 Mus muscu
72		16.8	80.0	1576	5	BX950346	BX950346 Gallus ga
73	c	16.8	80.0	4952	1	MVVHUGAUB	X61204 M.voltae vh
74	c	16.8	80.0	9431	9	BX571808	BX571808 Human DNA
75		16.8	80.0	29364	3	CEC27B7	Z54236 Caenorhabdi
76		16.8	80.0	39872	9	HSICB2046	Z97183 Human DNA s
77	c	16.8	80.0	69184	9	CR759793	CR759793 Human DNA
78		16.8	80.0	75518	2	AC100509	AC100509 Mus muscu
79	c	16.8	80.0	75518	2	AC100509	AC100509 Mus muscu
80	c	16.8	80.0	78635	9	AL662820	AL662820 Human DNA
81		16.8	80.0	79175	9	AC012516	AC012516 Homo sapi
82		16.8	80.0	83962	2	AP000448	AP000448 Homo sapi
83		16.8	80.0	85975	9	AC110011	AC110011 Homo sapi
84	c	16.8	80.0	86220	2	RN532D24	AL732655 Rattus no
85		16.8	80.0	94739	9	AP001964	AP001964 Homo sapi
86	c	16.8	80.0	108902	9	CR759817	CR759817 Human DNA
87		16.8	80.0	110000	2	AC103177_2	Continuation (3 of
88	c	16.8	80.0	110000	2	AC129386_1	Continuation (2 of
89		16.8	80.0	110000	2	AC145943_2	Continuation (3 of
90	c	16.8	80.0	110794	9	AL662827	AL662827 Human DNA
91		16.8	80.0	112209	10	AC022298	AC022298 Mus muscu
92	c	16.8	80.0	114575	9	BX248088	BX248088 Human DNA

	93	16.8	80.0	116695	2	CR762434	CR762434 Homo sapi
	94	16.8	80.0	117417	9	AL353704	AL353704 Human DNA
	95	16.8	80.0	128293	10	AL928607	AL928607 Mouse DNA
c	96	16.8	80.0	139807	2	AC105325	AC105325 Mus muscu
c	97	16.8	80.0	153700	2	AC118114	AC118114 Rattus no
c	98	16.8	80.0	159703	10	AC122439	AC122439 Mus muscu
	99	16.8	80.0	163758	2	AC141345	AC141345 Rattus no
	100	16.8	80.0	166305	2	AC022580	AC022580 Homo sapi

ALIGNMENTS

RESULT 1
AC109544
LOCUS AC109544 232064 bp DNA linear HTG 15-NOV-2002
DEFINITION Rattus norvegicus clone CH230-202O10, *** SEQUENCING IN PROGRESS ***.
ACCESSION AC109544
VERSION AC109544.5 GI:25006749
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 232064)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 232064)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (05-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 232064)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Nov 15, 2002 this sequence version replaced gi:23266105.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GQBR
Center clone name: CH230-202O10
---------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 210440 bases at least Q40
Consensus quality: 213054 bases at least Q30
Consensus quality: 214694 bases at least Q20
Estimated insert size: 221175; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 232064: contig of 232064 bp in length.

FEATURES Location/Qualifiers
source 1. .232064
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-202O10"
misc_feature 1. .1859
/note="wgs_end_extension
clone_end:T7"
misc_feature 4428. .5250
/note="clone_boundary
clone_end:T7
site:
end_sequence:BH335244"
misc_feature complement(230334. .231111)

```
                     /note="clone_boundary
                     clone_end:Sp6
                     site:
                     end_sequence:BH335245"
ORIGIN

  Query Match             92.4%;  Score 19.4;  DB 2;  Length 232064;
  Best Local Similarity   95.2%;  Pred. No. 63;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              ||||||||||||||||||  ||
Db      34510 CAGTGACATGCAGGTCTATCT 34530


RESULT 2
AC133226/c
LOCUS       AC133226               258823 bp    DNA     linear   HTG 15-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-329C22, *** SEQUENCING IN PROGRESS
            ***.
ACCESSION   AC133226
VERSION     AC133226.3  GI:25007420
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 258823)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 258823)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 258823)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 15, 2002 this sequence version replaced gi:22771302.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: KBNW
                Center clone name: CH230-329C22
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 187552 bases at least Q40
                Consensus quality: 190690 bases at least Q30
                Consensus quality: 192289 bases at least Q20
                Estimated insert size: 192259; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   258823: contig of 258823 bp in length.
FEATURES             Location/Qualifiers
     source          1. .258823
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-329C22"
     misc_feature    1. .1057
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    5970. .6869
                     /note="clone_boundary
                     clone_end:Sp6
                     site:
```

```
                     end_sequence:BZ184074"
     misc_feature    114027. .115223
                     /note="wgs_contig"
     misc_feature    140513. .142294
                     /note="wgs_contig"
     misc_feature    147730. .149278
                     /note="wgs_contig"
     misc_feature    199578. .200485
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:BZ184073"
     misc_feature    201856. .202987
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    206294. .207865
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    257672. .258823
                     /note="wgs_end_extension
                     clone_end:T7"
ORIGIN

  Query Match             92.4%;  Score 19.4;  DB 2;  Length 258823;
  Best Local Similarity   95.2%;  Pred. No. 62;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              |||||||||||||||||| ||
Db      21404 CAGTGACATGCAGGTCTATCT 21384


RESULT 3
CR352267/c
LOCUS       CR352267              140394 bp    DNA     linear   HTG 12-MAR-2004
DEFINITION  Danio rerio clone DKEY-174N5, WORKING DRAFT SEQUENCE, 9 unordered
            pieces.
ACCESSION   CR352267
VERSION     CR352267.4  GI:45433391
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 140394)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Mar 13, 2004 this sequence version replaced gi:45381849.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK174N5
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 138030 bases at least Q40
            Consensus quality: 138659 bases at least Q30
            Consensus quality: 139080 bases at least Q20
            Insert size: 139594; sum-of-contigs
            Insert size: 156923; 6.3% error; agarose-fp
            Quality coverage: 10.76x in Q20 bases; sum-of-contigs Quality
            coverage: 10.22x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     4851: contig of 4851 bp in length
            *     4852     4951: gap of 100 bp
            *     4952    15778: contig of 10827 bp in length
            *    15779    15878: gap of 100 bp
            *    15879    38168: contig of 22290 bp in length
            *    38169    38268: gap of 100 bp
            *    38269    54580: contig of 16312 bp in length
            *    54581    54680: gap of 100 bp
            *    54681    87936: contig of 33256 bp in length
            *    87937    88036: gap of 100 bp
            *    88037    94451: contig of 6415 bp in length
            *    94452    94551: gap of 100 bp
            *    94552   107512: contig of 12961 bp in length
            *   107513   107612: gap of 100 bp
            *   107613   112849: contig of 5237 bp in length
            *   112850   112949: gap of 100 bp
            *   112950   140394: contig of 27445 bp in length.
FEATURES             Location/Qualifiers
     source          1. .140394
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-174N5"
                     /clone_lib="DanioKey"
     misc_feature    1. .4851
                     /note="assembly_fragment:00165
                     fragment_chain:1"
     misc_feature    4952. .15778
                     /note="assembly_fragment:00354
                     fragment_chain:1"
     misc_feature    15879. .38168
                     /note="assembly_fragment:01046
                     fragment_chain:1"
     misc_feature    38269. .54580
                     /note="assembly_fragment:00715
                     fragment_chain:1"
     misc_feature    54681. .87936
                     /note="assembly_fragment:01919
                     fragment_chain:1"
     misc_feature    88037. .94451
                     /note="assembly_fragment:00241
                     fragment_chain:1"
     misc_feature    94552. .107512
                     /note="assembly_fragment:00500
                     fragment_chain:1"
     misc_feature    107613. .112849
                     /note="assembly_fragment:00118
                     fragment_chain:2"
     misc_feature    112950. .140394
                     /note="assembly_fragment:01411
                     fragment_chain:2"
ORIGIN

  Query Match             87.6%;  Score 18.4;  DB 2;  Length 140394;
  Best Local Similarity   95.0%;  Pred. No. 2.1e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 AGTGACATGCAGGTCTAGCT 21
              |||||| |||||||||||||
Db      74881 AGTGACCTGCAGGTCTAGCT 74862


RESULT 4
AP000772/c
LOCUS       AP000772              155858 bp    DNA     linear   HTG 30-MAY-2000
DEFINITION  Homo sapiens chromosome 11 clone CMB9-7B14 map 11q22, WORKING DRAFT
            SEQUENCE, 27 unordered pieces.
ACCESSION   AP000772
```

```
VERSION     AP000772.2  GI:8118931
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 155858)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 155,858 genomic DNA of 11q22
  JOURNAL   Published Only in DataBase (1999)
REFERENCE   2  (bases 1 to 155858)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-NOV-1999) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     On May 31, 2000 this sequence version replaced gi:6997610.
            -------------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
                Center project name: HumDraft11
                Center clone name: CMB9-7B14
            -------------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 135660 bases at least Q40
                Consensus quality: 145354 bases at least Q30
                Consensus quality: 150658 bases at least Q20
                Insert size: 153258; sum-of-contigs
                Quality coverage: 4.31x in Q20 bases; sum-of-contigs
            --------------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            27 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                    1     18489 contig of  18489 bp in length
                18590     36142 contig of  17553 bp in length
                36243     49897 contig of  13655 bp in length
                49998     63723 contig of  13726 bp in length
                63824     76791 contig of  12968 bp in length
                76892     85927 contig of   9036 bp in length
                86028     93056 contig of   7029 bp in length
                93157     99901 contig of   6745 bp in length
               100002    106585 contig of   6584 bp in length
               106686    111187 contig of   4502 bp in length
               111288    115206 contig of   3919 bp in length
               115307    119877 contig of   4571 bp in length
               119978    124080 contig of   4103 bp in length
               124181    127015 contig of   2835 bp in length
               127116    129487 contig of   2372 bp in length
               129588    133023 contig of   3436 bp in length
               133124    136332 contig of   3209 bp in length
               136433    139684 contig of   3252 bp in length
               139785    142916 contig of   3132 bp in length
               143017    145169 contig of   2153 bp in length
               145270    147638 contig of   2369 bp in length
               147739    148903 contig of   1165 bp in length
               149004    150381 contig of   1378 bp in length
               150482    151677 contig of   1196 bp in length
               151778    153587 contig of   1810 bp in length
               153688    154698 contig of   1011 bp in length
               154799    155858 contig of   1060 bp in length
            Sequence updated (26-May-2000).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 27 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    18489: contig of 18489 bp in length
            *   18490    18589: gap of 100 bp
            *   18590    36142: contig of 17553 bp in length
            *   36143    36242: gap of 100 bp
            *   36243    49897: contig of 13655 bp in length
            *   49898    49997: gap of 100 bp
            *   49998    63723: contig of 13726 bp in length
            *   63724    63823: gap of 100 bp
            *   63824    76791: contig of 12968 bp in length
            *   76792    76891: gap of 100 bp
            *   76892    85927: contig of 9036 bp in length
            *   85928    86027: gap of 100 bp
            *   86028    93056: contig of 7029 bp in length
            *   93057    93156: gap of 100 bp
            *   93157    99901: contig of 6745 bp in length
            *   99902   100001: gap of 100 bp
            *  100002   106585: contig of 6584 bp in length
            *  106586   106685: gap of 100 bp
            *  106686   111187: contig of 4502 bp in length
            *  111188   111287: gap of 100 bp
            *  111288   115206: contig of 3919 bp in length
            *  115207   115306: gap of 100 bp
            *  115307   119877: contig of 4571 bp in length
            *  119878   119977: gap of 100 bp
            *  119978   124080: contig of 4103 bp in length
            *  124081   124180: gap of 100 bp
            *  124181   127015: contig of 2835 bp in length
            *  127016   127115: gap of 100 bp
            *  127116   129487: contig of 2372 bp in length
            *  129488   129587: gap of 100 bp
            *  129588   133023: contig of 3436 bp in length
            *  133024   133123: gap of 100 bp
            *  133124   136332: contig of 3209 bp in length
            *  136333   136432: gap of 100 bp
            *  136433   139684: contig of 3252 bp in length
            *  139685   139784: gap of 100 bp
            *  139785   142916: contig of 3132 bp in length
            *  142917   143016: gap of 100 bp
            *  143017   145169: contig of 2153 bp in length
            *  145170   145269: gap of 100 bp
            *  145270   147638: contig of 2369 bp in length
            *  147639   147738: gap of 100 bp
            *  147739   148903: contig of 1165 bp in length
            *  148904   149003: gap of 100 bp
            *  149004   150381: contig of 1378 bp in length
            *  150382   150481: gap of 100 bp
            *  150482   151677: contig of 1196 bp in length
            *  151678   151777: gap of 100 bp
            *  151778   153587: contig of 1810 bp in length
            *  153588   153687: gap of 100 bp
            *  153688   154698: contig of 1011 bp in length
            *  154699   154798: gap of 100 bp
            *  154799   155858: contig of 1060 bp in length.
FEATURES             Location/Qualifiers
     source          1. .155858
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q22"
                     /clone="CMB9-7B14"
     misc_feature    1. .18489
                     /note="assembly_fragment"
```

```
     misc_feature    18590. .36142
                     /note="assembly_fragment"
     misc_feature    36243. .49897
                     /note="assembly_fragment"
     misc_feature    49998. .63723
                     /note="assembly_fragment"
     misc_feature    63824. .76791
                     /note="assembly_fragment"
     misc_feature    76892. .85927
                     /note="assembly_fragment"
     misc_feature    86028. .93056
                     /note="assembly_fragment"
     misc_feature    93157. .99901
                     /note="assembly_fragment"
     misc_feature    100002. .106585
                     /note="assembly_fragment"
     misc_feature    106686. .111187
                     /note="assembly_fragment"
     misc_feature    111288. .115206
                     /note="assembly_fragment"
     misc_feature    115307. .119877
                     /note="assembly_fragment"
     misc_feature    119978. .124080
                     /note="assembly_fragment"
     misc_feature    124181. .127015
                     /note="assembly_fragment"
     misc_feature    127116. .129487
                     /note="assembly_fragment"
     misc_feature    129588. .133023
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
     misc_feature    133124. .136332
                     /note="assembly_fragment"
     misc_feature    136433. .139684
                     /note="assembly_fragment"
     misc_feature    139785. .142916
                     /note="assembly_fragment"
     misc_feature    143017. .145169
                     /note="assembly_fragment"
     misc_feature    145270. .147638
                     /note="assembly_fragment"
     misc_feature    147739. .148903
                     /note="assembly_fragment"
     misc_feature    149004. .150381
                     /note="assembly_fragment"
     misc_feature    150482. .151677
                     /note="assembly_fragment"
     misc_feature    151778. .153587
                     /note="assembly_fragment"
     misc_feature    153688. .154698
                     /note="assembly_fragment"

  Query Match             87.6%;  Score 18.4;  DB 2;  Length 155858;
  Best Local Similarity   95.0%;  Pred. No. 2.1e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGC 20
              |||| ||||||||||||||||
Db      89083 CAGTTACATGCAGGTCTAGC 89064


RESULT 5
CR391906/c
LOCUS       CR391906              161516 bp    DNA     linear   HTG 24-APR-2004
DEFINITION  Danio rerio clone DKEY-211K10, *** SEQUENCING IN PROGRESS ***, 9
            unordered pieces.
ACCESSION   CR391906
VERSION     CR391906.2  GI:46559615
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 161516)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 24, 2004 this sequence version replaced gi:46517964.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK211K10
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 159202 bases at least Q40
            Consensus quality: 159648 bases at least Q30
            Consensus quality: 160053 bases at least Q20
            Insert size: 160716; sum-of-contigs
            Insert size: 167493; 3.3% error; agarose-fp
            Quality coverage: 6.27x in Q20 bases; sum-of-contigs Quality
            coverage: 6.01x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     2483: contig of 2483 bp in length
            *    2484     2583: gap of 100 bp
            *    2584    38899: contig of 36316 bp in length
            *   38900    38999: gap of 100 bp
            *   39000    55041: contig of 16042 bp in length
            *   55042    55141: gap of 100 bp
            *   55142    58203: contig of 3062 bp in length
            *   58204    58303: gap of 100 bp
            *   58304    70412: contig of 12109 bp in length
            *   70413    70512: gap of 100 bp
            *   70513   100144: contig of 29632 bp in length
            *  100145   100244: gap of 100 bp
            *  100245   110317: contig of 10073 bp in length
            *  110318   110417: gap of 100 bp
            *  110418   122844: contig of 12427 bp in length
            *  122845   122944: gap of 100 bp
            *  122945   161516: contig of 38572 bp in length.
FEATURES             Location/Qualifiers
     source          1. .161516
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-211K10"
                     /clone_lib="DanioKey"
     misc_feature    1. .2483
                     /note="assembly_fragment:00070
                     fragment_chain:1"
     misc_feature    2584. .38899
                     /note="assembly_fragment:01225
                     fragment_chain:1"
     misc_feature    39000. .55041
                     /note="assembly_fragment:00464
                     fragment_chain:1"
     misc_feature    55142. .58203
                     /note="assembly_fragment:00087
                     fragment_chain:1"
     misc_feature    58304. .70412
                     /note="assembly_fragment:00114
                     fragment_chain:2"
     misc_feature    70513. .100144
```

```
                     /note="assembly_fragment:00640
                     fragment_chain:2"
     misc_feature    100245. .110317
                     /note="assembly_fragment:00215
                     fragment_chain:3"
     misc_feature    110418. .122844
                     /note="assembly_fragment:00326
                     fragment_chain:3"
     misc_feature    122945. .161516
                     /note="assembly_fragment:00891.0"
ORIGIN

  Query Match             87.6%;  Score 18.4;  DB 2;  Length 161516;
  Best Local Similarity   95.0%;  Pred. No. 2.1e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 AGTGACATGCAGGTCTAGCT 21
              |||||| ||||||||||||||
Db     141651 AGTGACCTGCAGGTCTAGCT 141632


RESULT 6
U21308
LOCUS       U21308                 40392 bp    DNA     linear   INV 29-SEP-2004
DEFINITION  Caenorhabditis elegans cosmid ZK1290, complete sequence.
ACCESSION   U21308
VERSION     U21308.1  GI:687795
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 40392)
  AUTHORS   .
  CONSRTM   WormBase Consortium
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
  MEDLINE   99069613
   PUBMED   9851916
REFERENCE   2  (bases 1 to 40392)
  AUTHORS   Taich,A.
  TITLE     The sequence of C. elegans cosmid ZK1290
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 40392)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-1995) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   4  (bases 1 to 40392)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1997) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   5  (bases 1 to 40392)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2001) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   6  (bases 1 to 40392)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAY-2002) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   7  (bases 1 to 40392)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-2002) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   8  (bases 1 to 40392)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2002) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   9  (bases 1 to 40392)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-DEC-2002) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   10 (bases 1 to 40392)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-2003) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   11 (bases 1 to 40392)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-2003) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   12 (bases 1 to 40392)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-NOV-2003) Department of Genetics, Washington
            University,.Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   13 (bases 1 to 40392)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-2003) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   14 (bases 1 to 40392)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-2004) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   15 (bases 1 to 40392)
  AUTHORS   .
  CONSRTM   WormBase Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (29-SEP-2004) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
COMMENT     Submitted by:
                  Genome Sequencing Center
                  Department of Genetics, Washington University
                  St. Louis , MO  63110, USA, and
                  Sanger Centre, Hinxton Hall
                  Cambridge CB10 IRQ, England
                  email: submissions@watson.wustl.edu and jes@sanger.ac.uk

            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate chemistry
            or covered by high quality data (i.e., phred quality >= 30); an
            attempt was made to resolve all sequencing problems, such as
            compressions and repeats; all regions were covered by sequence from
            more than one m13 subclone.

            For a graphical representation of this clone sequence and its
            analysis see:
            http://www.wormbase.org/db/seq/sequence?name=ZK1290;class=Sequence
```

NEIGHBORING CLONE INFORMATION

The 5' clone is C28F5, 200 bp overlap; the 3' clone is F35D2, 200 bp overlap. Actual start of this clone is at base position 1 of ZK1290; actual end is at 40392 of ZK1290.

NOTES:

Coding seqences below are the result of integration and manual review of the following data : computer analysis using the program Genefinder (P. Green and L. Hillier, personal communication), the large scale EST projects of Yuji Kohara (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C. elegans ORFeome cloning project (http://worfdb.dfci.harvard.edu/), similarity to other proteins from BlastX analyses (http://blast.wustl.edu/), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans GenBank submissions, and personal communications with C. elegans researchers. tRNAs are predicted using the program tRNAscan-SE (Lowe, T.M. and Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

```
FEATURES             Location/Qualifiers
     source          1. .40392
                     /organism="Caenorhabditis elegans"
                     /mol_type="genomic DNA"
                     /strain="Bristol N2"
                     /db_xref="taxon:6239"
                     /chromosome="II"
                     /clone="ZK1290"
     gene            complement(1177. .2162)
                     /gene="ZK1290.7"
                     /locus_tag="ZK1290.7"
     CDS             complement(join(1177. .1344,1389. .1499,1544. .1623,
                     1668. .2043,2085. .2162))
                     /gene="ZK1290.7"
                     /locus_tag="ZK1290.7"
                     /standard_name="ZK1290.7"
                     /note="contains similarity to Pfam domain PF01482 (Domain
                     of unknown function DUF13)"
                     /codon_start=1
                     /product="Hypothetical protein ZK1290.7"
                     /protein_id="AAB93314.2"
                     /db_xref="GI:38176088"
                     /db_xref="WormBase:ZK1290.7"
                     /translation="MTAYAKWSSCFQNGIDSWNNLTTTAEFLNHAGNISSNCKIILDL
                     NIVKLSDSKFYISPQVQKSLTHATVGISSNFEAESQLKKVLELSSHHFLICQLFQRLP
                     IINIFGATPKLSERERELFQKIGNYYPVAIGTKSGLVNVTMKPQYNYSLNDSAVPSSK
                     SEFQLDLIYFLKDVLKQDFYYSIWFNIFFDELPNMLHRNRNFDENRITLCQININIDN
                     DQKTNMKQFFKSLISDKRYVIIFEVVNDYGVLFLINFDNTTCRKKLGIVRNV"
     gene            2459. .4869
                     /gene="ZK1290.6"
                     /locus_tag="ZK1290.6"
     CDS             join(2459. .2528,2570. .2601,2649. .2691,2741. .2899,
                     3014. .3079,3124. .3324,3378. .3425,3468. .3554,3599. .3655,
                     3699. .3740,3791. .3855,3902. .4015,4060. .4241,4378. .4449,
                     4494. .4599,4648. .4747,4790. .4869)
                     /gene="ZK1290.6"
                     /locus_tag="ZK1290.6"
                     /standard_name="ZK1290.6"
                     /note="contains similarity to Pfam domain PF00075 (RNase
                     H); coded for by the following C. elegans cDNAs:
                     OSTR116H2_1, yk367d2.3, yk367d2.5, yk1479d07.3,
                     yk1479d07.5"
                     /codon_start=1
                     /product="Hypothetical protein ZK1290.6"
                     /protein_id="AAB93313.3"
                     /db_xref="GI:38176087"
                     /db_xref="WormBase:ZK1290.6"
                     /translation="MIRWFRNFGALFKKPRGAGLMKNEQTHNLQRVVQNDITCPEEQR
                     STGGKGGGGGSHGGHGGSHGGSHGGHGSHGGSYGGHSGHGRSFSTSSGSSGYRGLSSG
                     STNTSYSKSTPRTNSGYSTGKTTKTASSSRNYSNTSSTRSTTKPSKSSYSKRSTKSTK
                     SLNYSSSSLGSKSIEKSRRSSRKSGSRQDSVASRRQLRSVSSKPQKKYVATDMPSTVS
                     TGRSRGRSVYRRNNVNADVSQRQSRSMSVSRGRSRNRSQNRERSPSIEKSTRSQHVSR
                     KDAFNSHDVLSPTNRSGSVHRSHSAHHERTFEERHKTPNGVWNEMFPKGTLVMYTDGS
                     YLKRPPTSGIGIFVGPGHELNRSQRIRGPIQDNNYAEFIAVRTALQNALKNENYRDQK
                     VVIRTDCLNVIEALQGTRPTAFVDVKSQVEFLSKQFPKGVHFQHVYAHAGDPGNEMAD
                     LFAGQASSKSQNSFGGGQRTRSASTARGRSRERNRLRSKSNDPSFIRSRSHSIHSRK"

  Query Match             84.8%;  Score 17.8;  DB 3;  Length 40392;
  Best Local Similarity   90.5%;  Pred. No. 4.8e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              |||| ||||||| ||||||||
Db      33102 CAGTTACATGCAAGTCTAGCT 33122


RESULT 7
AC100344/c
LOCUS       AC100344               72655 bp    DNA     linear   HTG 22-NOV-2001
DEFINITION  Mus musculus clone RP23-127F18, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC100344
VERSION     AC100344.1  GI:17047710
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 72655)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-127F18
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 72655)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L15140
                Center clone name: 127_F_18
            --------------.
            * NOTE: This record contains 67 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
```

```
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*        1      949: contig of 949 bp in length
*      950     1049: gap of 100 bp
*     1050     2006: contig of 957 bp in length
*     2007     2106: gap of 100 bp
*     2107     3119: contig of 1013 bp in length
*     3120     3219: gap of 100 bp
*     3220     4172: contig of 953 bp in length
*     4173     4272: gap of 100 bp
*     4273     5244: contig of 972 bp in length
*     5245     5344: gap of 100 bp
*     5345     6288: contig of 944 bp in length
*     6289     6388: gap of 100 bp
*     6389     7363: contig of 975 bp in length
*     7364     7463: gap of 100 bp
*     7464     8447: contig of 984 bp in length
*     8448     8547: gap of 100 bp
*     8548     9561: contig of 1014 bp in length
*     9562     9661: gap of 100 bp
*     9662    10597: contig of 936 bp in length
*    10598    10697: gap of 100 bp
*    10698    11718: contig of 1021 bp in length
*    11719    11818: gap of 100 bp
*    11819    12831: contig of 1013 bp in length
*    12832    12931: gap of 100 bp
*    12932    13900: contig of 969 bp in length
*    13901    14000: gap of 100 bp
*    14001    14953: contig of 953 bp in length
*    14954    15053: gap of 100 bp
*    15054    16082: contig of 1029 bp in length
*    16083    16182: gap of 100 bp
*    16183    17195: contig of 1013 bp in length
*    17196    17295: gap of 100 bp
*    17296    18344: contig of 1049 bp in length
*    18345    18444: gap of 100 bp
*    18445    19385: contig of 941 bp in length
*    19386    19485: gap of 100 bp
*    19486    20455: contig of 970 bp in length
*    20456    20555: gap of 100 bp
*    20556    21548: contig of 993 bp in length
*    21549    21648: gap of 100 bp
*    21649    22631: contig of 983 bp in length
*    22632    22731: gap of 100 bp
*    22732    23705: contig of 974 bp in length
*    23706    23805: gap of 100 bp
*    23806    24762: contig of 957 bp in length
*    24763    24862: gap of 100 bp
*    24863    25835: contig of 973 bp in length
*    25836    25935: gap of 100 bp
*    25936    26918: contig of 983 bp in length
*    26919    27018: gap of 100 bp
*    27019    28003: contig of 985 bp in length
*    28004    28103: gap of 100 bp
*    28104    29109: contig of 1006 bp in length
*    29110    29209: gap of 100 bp
*    29210    30234: contig of 1025 bp in length
*    30235    30334: gap of 100 bp
*    30335    31327: contig of 993 bp in length
*    31328    31427: gap of 100 bp
*    31428    32398: contig of 971 bp in length
*    32399    32498: gap of 100 bp
*    32499    33501: contig of 1003 bp in length
*    33502    33601: gap of 100 bp
*    33602    34612: contig of 1011 bp in length
*    34613    34712: gap of 100 bp
*    34713    35718: contig of 1006 bp in length
*    35719    35818: gap of 100 bp
*    35819    36820: contig of 1002 bp in length
*    36821    36920: gap of 100 bp
*    36921    37914: contig of 994 bp in length
*    37915    38014: gap of 100 bp
*    38015    38989: contig of 975 bp in length
*    38990    39089: gap of 100 bp
*    39090    40045: contig of 956 bp in length
*    40046    40145: gap of 100 bp
*    40146    41107: contig of 962 bp in length
*    41108    41207: gap of 100 bp
*    41208    42219: contig of 1012 bp in length
*    42220    42319: gap of 100 bp
*    42320    43325: contig of 1006 bp in length
*    43326    43425: gap of 100 bp
*    43426    44439: contig of 1014 bp in length
*    44440    44539: gap of 100 bp
*    44540    45494: contig of 955 bp in length
*    45495    45594: gap of 100 bp
*    45595    46580: contig of 986 bp in length
*    46581    46680: gap of 100 bp
*    46681    47695: contig of 1015 bp in length
*    47696    47795: gap of 100 bp
*    47796    48789: contig of 994 bp in length
*    48790    48889: gap of 100 bp
*    48890    49831: contig of 942 bp in length
*    49832    49931: gap of 100 bp
*    49932    50926: contig of 995 bp in length
*    50927    51026: gap of 100 bp
*    51027    52016: contig of 990 bp in length
*    52017    52116: gap of 100 bp
*    52117    53074: contig of 958 bp in length
*    53075    53174: gap of 100 bp
*    53175    54218: contig of 1044 bp in length
*    54219    54318: gap of 100 bp
*    54319    55335: contig of 1017 bp in length
*    55336    55435: gap of 100 bp
*    55436    56368: contig of 933 bp in length
*    56369    56468: gap of 100 bp
*    56469    57441: contig of 973 bp in length
*    57442    57541: gap of 100 bp
*    57542    58519: contig of 978 bp in length
*    58520    58619: gap of 100 bp
*    58620    59585: contig of 966 bp in length
*    59586    59685: gap of 100 bp
*    59686    60652: contig of 967 bp in length
*    60653    60752: gap of 100 bp
*    60753    61736: contig of 984 bp in length
*    61737    61836: gap of 100 bp
*    61837    62768: contig of 932 bp in length
*    62769    62868: gap of 100 bp
*    62869    63880: contig of 1012 bp in length
*    63881    63980: gap of 100 bp
*    63981    64974: contig of 994 bp in length
*    64975    65074: gap of 100 bp
*    65075    66085: contig of 1011 bp in length
*    66086    66185: gap of 100 bp
*    66186    67121: contig of 936 bp in length
*    67122    67221: gap of 100 bp
*    67222    68235: contig of 1014 bp in length
*    68236    68335: gap of 100 bp
*    68336    69321: contig of 986 bp in length
*    69322    69421: gap of 100 bp
*    69422    70456: contig of 1035 bp in length
*    70457    70556: gap of 100 bp
*    70557    71570: contig of 1014 bp in length
*    71571    71670: gap of 100 bp
*    71671    72655: contig of 985 bp in length.
FEATURES             Location/Qualifiers
     source          1. .72655
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"

  Query Match             84.8%;  Score 17.8;  DB 2;  Length 72655;
  Best Local Similarity   90.5%;  Pred. No. 4.5e+02;
```

```
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              | |||||||||||| |||||||
Db      26318 CTGTGACATGCAGATCTAGCT 26298


RESULT 8
AL831718/c
LOCUS       AL831718              116076 bp    DNA     linear   ROD 15-NOV-2002
DEFINITION  Mouse DNA sequence from clone RP23-146O20 on chromosome X, complete
            sequence.
ACCESSION   AL831718
VERSION     AL831718.6  GI:25137019
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 116076)
  AUTHORS   Clark,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 19, 2002 this sequence version replaced gi:22213735.
            ------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-146O20 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .116076
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="X"
                     /clone="RP23-146O20"
                     /clone_lib="RPCI-23"
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 10;  Length 116076;
  Best Local Similarity   90.5%;  Pred. No. 4.2e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              | |||||||||||| |||||||
Db     107376 CTGTGACATGCAGATCTAGCT 107356


RESULT 9
AC148761/c
LOCUS       AC148761              125354 bp    DNA     linear   HTG 02-APR-2004
DEFINITION  Medicago truncatula chromosome 2 clone mth2-19h23, *** SEQUENCING
            IN PROGRESS ***, 14 unordered pieces.
ACCESSION   AC148761
VERSION     AC148761.1  GI:46063628
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 125354)
  AUTHORS   Town,C.D., Tallon,L.J., Arbogast,T., Althoff,R., Hine,E.,
            Monaghan,E., Smith,S.A., Utterback,T., Feldblyum,T. and Koo,H.
  TITLE     Medicago truncatula BAC genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 125354)
  AUTHORS   Town,C.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-APR-2004) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1075: contig of 1075 bp in length
            *     1076     1175: gap of unknown length
            *     1176     2410: contig of 1235 bp in length
            *     2411     2510: gap of unknown length
            *     2511    29548: contig of 27038 bp in length
            *    29549    29648: gap of unknown length
            *    29649    32369: contig of 2721 bp in length
            *    32370    32469: gap of unknown length
            *    32470    44227: contig of 11758 bp in length
            *    44228    44327: gap of unknown length
            *    44328    45752: contig of 1425 bp in length
            *    45753    45852: gap of unknown length
            *    45853    77661: contig of 31809 bp in length
            *    77662    77761: gap of unknown length
            *    77762    79043: contig of 1282 bp in length
            *    79044    79143: gap of unknown length
            *    79144    88662: contig of 9519 bp in length
            *    88663    88762: gap of unknown length
            *    88763    91773: contig of 3011 bp in length
            *    91774    91873: gap of unknown length
            *    91874    95698: contig of 3825 bp in length
            *    95699    95798: gap of unknown length
            *    95799   103010: contig of 7212 bp in length
            *   103011   103110: gap of unknown length
            *   103111   111029: contig of 7919 bp in length
            *   111030   111129: gap of unknown length
            *   111130   125354: contig of 14225 bp in length.
FEATURES             Location/Qualifiers
     source          1. .125354
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3880"
                     /chromosome="2"
                     /clone="mth2-19h23"
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 2;  Length 125354;
  Best Local Similarity   90.5%;  Pred. No. 4.2e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
```

```
                ||||||||||||||| ||| ||
Db      42736 CAGTGACATGCAGGCCTACCT 42716

RESULT 10
AC110379
LOCUS       AC110379               139884 bp    DNA     linear   ROD 05-NOV-2003
DEFINITION  Mus musculus BAC clone RP24-193N24 from 15, complete sequence.
ACCESSION   AC110379
VERSION     AC110379.3  GI:21536156
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 139884)
  AUTHORS   Haakenson,W. and Doebber,A.
  TITLE     The sequence of Mus musculus BAC clone RP24-193N24
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 139884)
  AUTHORS   Wilson,R.
  TITLE     Sequencing of Mus musculus
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 139884)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-FEB-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 139884)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   5  (bases 1 to 139884)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUN-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   6  (bases 1 to 139884)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-NOV-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Jun 21, 2002 this sequence version replaced gi:20069820.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: M_BB0193N24
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-24 BAC Library has been constructed by Pieter de Jong and
            coworkers (http://www.chori.org) from male C57BL/6J mouse spleen
            and/or brain genomic DNA. The clone and detailed information can be
            obtained from Pieter de Jong and coworkers at http://www.chori.org

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is the entire insert of the clone.
FEATURES             Location/Qualifiers
     source          1. .139884
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="15"
                     /map="15"
                     /clone="RP24-193N24"
                     /clone_lib="RPCI-24"
     repeat_region   1472. .1532
                     /rpt_family="Alu"
     repeat_region   2336. .2478
                     /rpt_family="B4"
     repeat_region   3625. .3939
                     /rpt_family="MaLR"
     repeat_region   4251. .4445
                     /rpt_family="B2"
     repeat_region   4623. .4984
                     /rpt_family="MaLR"
     repeat_region   6912. .7089
                     /rpt_family="B2"
     repeat_region   7105. .7222
                     /rpt_family="B4"
     repeat_region   7264. .7331
                     /rpt_family="Alu"
     repeat_region   7640. .7961
                     /rpt_family="L1"
     repeat_region   9097. .9389
                     /rpt_family="B4"
     repeat_region   11045. .11288
                     /rpt_family="MaLR"
     repeat_region   11701. .11779
                     /rpt_family="L1"
     repeat_region   12001. .12197
                     /rpt_family="B2"
     repeat_region   12903. .13063
                     /rpt_family="MER1_type"
     repeat_region   13539. .13735
                     /rpt_family="B2"
     repeat_region   14427. .14547
                     /rpt_family="B2"
     repeat_region   15060. .15237
                     /rpt_family="L1"
     repeat_region   15638. .15782
                     /rpt_family="Alu"
     repeat_region   16006. .16312
                     /rpt_family="B4"
     repeat_region   17607. .17890
                     /rpt_family="MaLR"
     repeat_region   18016. .18186
                     /rpt_family="L1"
     repeat_region   18215. .18343
                     /rpt_family="L1"
     repeat_region   19190. .19288
                     /rpt_family="B4"
     repeat_region   19289. .19567
                     /rpt_family="MaLR"
     repeat_region   19574. .19671
                     /rpt_family="B4"
     repeat_region   21199. .21262
                     /rpt_family="4.5SRNA"
     repeat_region   21210. .21311
                     /rpt_family="Alu"
     repeat_region   21500. .22334
                     /rpt_family="L1"
     repeat_region   22533. .22620
```

```
                     /rpt_family="B4"
     repeat_region   23026. .23125
                     /rpt_family="Alu"
     repeat_region   23974. .24135
                     /rpt_family="B4"
     repeat_region   24176. .24297
                     /rpt_family="Alu"
     repeat_region   24344. .24414
                     /rpt_family="ID"
     repeat_region   24809. .24900
                     /rpt_family="L1"
     repeat_region   25063. .25207
                     /rpt_family="MaLR"
     repeat_region   26971. .27396
                     /rpt_family="MaLR"
     repeat_region   27790. .28076
                     /rpt_family="MaLR"
     repeat_region   28465. .28682
                     /rpt_family="B2"
     repeat_region   28746. .28932
                     /rpt_family="B4"
     repeat_region   29519. .29653
                     /rpt_family="B4"
     repeat_region   29622. .29899
                     /rpt_family="MaLR"
     repeat_region   30110. .30243
                     /rpt_family="MER1_type"
     repeat_region   31433. .31800
                     /rpt_family="MaLR"
     repeat_region   32501. .32551
                     /rpt_family="L2"
     repeat_region   32674. .32828
                     /rpt_family="MIR"
     repeat_region   34609. .34755
                     /rpt_family="Alu"
     repeat_region   35577. .35732
                     /rpt_family="B4"
     repeat_region   36028. .36213
                     /rpt_family="B2"
     repeat_region   37888. .38130
                     /rpt_family="B4"
     repeat_region   38777. .38983
                     /rpt_family="B4"
     repeat_region   39025. .39207
                     /rpt_family="B2"
     repeat_region   40928. .41260
                     /rpt_family="MaLR"
     repeat_region   41364. .41539
                     /rpt_family="B2"
     repeat_region   42540. .42702
                     /rpt_family="MER1_type"
     repeat_region   43219. .43441
                     /rpt_family="B4"
     repeat_region   43887. .44000
                     /rpt_family="B4"
     repeat_region   44624. .44961
                     /rpt_family="MaLR"
     repeat_region   45125. .45221
                     /rpt_family="MER2_type"
     repeat_region   46183. .46321

  Query Match             84.8%;  Score 17.8;  DB 10;  Length 139884;
  Best Local Similarity   90.5%;  Pred. No. 4.1e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              ||||||| |||||| ||||||
Db       9945 CAGTGACTTGCAGGGCTAGCT 9965


RESULT 11
AC102916/c
LOCUS       AC102916               144142 bp    DNA     linear   ROD 29-SEP-2004
```

```
DEFINITION  Mus musculus chromosome 5, clone RP24-274I18, complete sequence.
ACCESSION   AC102916
VERSION     AC102916.5  GI:52839774
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 144142)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 5, clone RP24-274I18
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 144142)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 144142)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
            Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J.,
            Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
            DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
            Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,
            Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
            Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
            MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
            Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
            Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-AUG-2004) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 144142)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
            Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J.,
            Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
            DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
            Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,
            Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
```

```
            Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
            MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
            Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
            Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-SEP-2004) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Sep 29, 2004 this sequence version replaced gi:51491651.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@broad.mit.edu
            -------------- Project Information
                Center project name: L20072
                Center clone name: 274_I_18
            --------------.
FEATURES             Location/Qualifiers
     source          1. .144142
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="5"
                     /map="5"
                     /clone="RP24-274I18"
                     /clone_lib="RPCI-24 Male Mouse BAC"
     repeat_region   1809. .1907
                     /rpt_family="L1MA7"
     repeat_region   3379. .3407
                     /rpt_family="AT_rich"
     repeat_region   4416. .4449
                     /rpt_family="(TCTA)n"
     repeat_region   4491. .4626
                     /rpt_family="MIR"
     repeat_region   4750. .4796
                     /rpt_family="(TTTTC)n"
     repeat_region   5588. .5972
                     /rpt_family="MTD"
     repeat_region   complement(6228. .6446)
                     /rpt_family="B4A"
     repeat_region   complement(6975. .7231)
                     /rpt_family="B4A"
     repeat_region   7508. .7533
                     /rpt_family="(GGGTG)n"
     repeat_region   complement(7620. .7829)
                     /rpt_family="B3"
     repeat_region   complement(8983. .9104)
                     /rpt_family="B1F"
     repeat_region   complement(9124. .9285)
                     /rpt_family="MIR"
     repeat_region   9796. .9816
                     /rpt_family="AT_rich"
     repeat_region   10137. .10168
                     /rpt_family="(CGTG)n"
     repeat_region   10547. .10581
                     /rpt_family="(TG)n"
     repeat_region   10584. .10610
                     /rpt_family="(CA)n"
     repeat_region   13802. .13832
                     /rpt_family="AT_rich"
     repeat_region   13872. .13907
                     /rpt_family="(TTTC)n"
     repeat_region   complement(13910. .13993)
                     /rpt_family="RMER17C"
     repeat_region   14015. .14062
                     /rpt_family="CT-rich"
     repeat_region   14076. .14101
                     /rpt_family="(A)n"
     repeat_region   complement(14102. .14205)
                     /rpt_family="PB1D9"
     repeat_region   14225. .14346
                     /rpt_family="(TTCC)n"
     unsure          complement(14259. .14376)
                     /note="single clone coverage"
     repeat_region   14348. .14392
                     /rpt_family="T-rich"
     repeat_region   14413. .14560
                     /rpt_family="B1_MM"
     repeat_region   15045. .15077
                     /rpt_family="(CAA)n"
     repeat_region   complement(15716. .16012)
                     /rpt_family="Lx9"
     repeat_region   16073. .16124
                     /rpt_family="(CA)n"
     repeat_region   18492. .18513
                     /rpt_family="(AGGGGG)n"
     repeat_region   20380. .20413
                     /rpt_family="AT_rich"
     repeat_region   21700. .21840
                     /rpt_family="B1_MM"
     repeat_region   21841. .21870
                     /rpt_family="(A)n"
     repeat_region   complement(22919. .22991)
                     /rpt_family="L3"
     repeat_region   23499. .23610
                     /rpt_family="L2"
     repeat_region   27822. .28224
                     /rpt_family="MTC"
     repeat_region   complement(28484. .29277)
                     /rpt_family="MT-int"
     repeat_region   complement(29293. .29404)
                     /rpt_family="MTD"
     repeat_region   complement(29396. .29529)
                     /rpt_family="B3"
     repeat_region   29660. .29776
                     /rpt_family="PB1D10"
     repeat_region   complement(29798. .30152)
                     /rpt_family="BGLII"
     repeat_region   complement(31140. .31406)
                     /rpt_family="Charlie1"
     repeat_region   complement(31556. .31742)
                     /rpt_family="B2_Mm2"
     repeat_region   31743. .31786
                     /rpt_family="(TTG)n"
     repeat_region   32140. .32160

  Query Match             84.8%;  Score 17.8;  DB 10;  Length 144142;
  Best Local Similarity   90.5%;  Pred. No. 4.1e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              ||| | ||||||||||||||||
Db      75702 CAGAGCCATGCAGGTCTAGCT 75682


RESULT 12
AC097230/c
LOCUS       AC097230              162289 bp    DNA     linear   MAM 28-JAN-2003
DEFINITION  Sus scrofa clone RP44-254G1, complete sequence.
ACCESSION   AC097230
VERSION     AC097230.3  GI:21956556
KEYWORDS    HTG.
SOURCE      Sus scrofa (pig)
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 162289)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
            Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P.,
            Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., Mastrian,S.D.,
            McCloskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,
            Portnoy,M.E., Prasad,A., Schueler,M.G., Stantripop,S., Thomas,J.W.,
            Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 162289)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-OCT-2001) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 162289)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   4  (bases 1 to 162289)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   5  (bases 1 to 162289)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JAN-2003) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Jul 26, 2002 this sequence version replaced gi:18860664.
            -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
                 Center project name: cnz
                 Center clone name: 254G01

            This sequence was finished as follows unless otherwise noted:
            all regions were double-stranded, sequenced with an
            alternate chemistry, or covered by high quality data
            (i.e., phred quality >= 30); an attempt was made to resolve
            all sequencing problems, such as compressions and repeats;
            all regions were covered by at least one plasmid subclone
            or more than one M13 subclone; and the assembly was confirmed
            by restriction digest.

            CLONE LENGTH: This sequence represents the entire insert of
            this clone unless otherwise noted. If there are overlapping
            clones, the overlaps are noted in the beginning and end of
            the Features section.
FEATURES             Location/Qualifiers
     source          1. .162289
                     /organism="Sus scrofa"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9823"
                     /clone="RP44-254G1"
                     /clone_lib="RP44"
     misc_feature    4934. .5067
                     /note="single clone coverage"
     misc_feature    49431. .49519
                     /note="single clone coverage"
     misc_feature    91219. .91237
                     /note="single clone coverage"
     misc_feature    150551. .150821
                     /note="single clone coverage"
     misc_feature    152274. .152290
                     /note="single clone coverage"
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 4;  Length 162289;
  Best Local Similarity   90.5%;  Pred. No. 4.1e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              ||||||||||| ||||| ||||
Db      32212 CAGTGACATGAAGGTCAAGCT 32192


RESULT 13
AC069005/c
LOCUS       AC069005              162560 bp    DNA     linear   HTG 16-JUL-2000
DEFINITION  Homo sapiens chromosome 8 clone RP11-712I15, WORKING DRAFT
            SEQUENCE, 33 unordered pieces.
ACCESSION   AC069005
VERSION     AC069005.3  GI:8844180
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 162560)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 162560)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jun 30, 2000 this sequence version replaced gi:8469066.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0712I15
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 145600 bases at least Q40
            Consensus quality: 151065 bases at least Q30
            Consensus quality: 153390 bases at least Q20
            Insert size: 173000; agarose-fp
            Insert size: 159360; sum-of-contigs
            Quality coverage: 3.35 in Q20 bases; agarose-fp
            Quality coverage: 3.73 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 33 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1483: contig of 1483 bp in length
            *     1484     1583: gap of unknown length
            *     1584     3194: contig of 1611 bp in length
            *     3195     3294: gap of unknown length
            *     3295     4472: contig of 1178 bp in length
            *     4473     4572: gap of unknown length
            *     4573     6120: contig of 1548 bp in length
            *     6121     6220: gap of unknown length
            *     6221     8046: contig of 1826 bp in length
```

```
            *     8047     8146: gap of unknown length
            *     8147     9901: contig of 1755 bp in length
            *     9902    10001: gap of unknown length
            *    10002    12500: contig of 2499 bp in length
            *    12501    12600: gap of unknown length
            *    12601    15266: contig of 2666 bp in length
            *    15267    15366: gap of unknown length
            *    15367    17549: contig of 2183 bp in length
            *    17550    17649: gap of unknown length
            *    17650    20378: contig of 2729 bp in length
            *    20379    20478: gap of unknown length
            *    20479    23345: contig of 2867 bp in length
            *    23346    23445: gap of unknown length
            *    23446    27122: contig of 3677 bp in length
            *    27123    27222: gap of unknown length
            *    27223    30048: contig of 2826 bp in length
            *    30049    30148: gap of unknown length
            *    30149    32220: contig of 2072 bp in length
            *    32221    32320: gap of unknown length
            *    32321    35967: contig of 3647 bp in length
            *    35968    36067: gap of unknown length
            *    36068    39730: contig of 3663 bp in length
            *    39731    39830: gap of unknown length
            *    39831    43447: contig of 3617 bp in length
            *    43448    43547: gap of unknown length
            *    43548    47467: contig of 3920 bp in length
            *    47468    47567: gap of unknown length
            *    47568    51880: contig of 4313 bp in length
            *    51881    51980: gap of unknown length
            *    51981    57315: contig of 5335 bp in length
            *    57316    57415: gap of unknown length
            *    57416    61833: contig of 4418 bp in length
            *    61834    61933: gap of unknown length
            *    61934    65951: contig of 4018 bp in length
            *    65952    66051: gap of unknown length
            *    66052    70869: contig of 4818 bp in length
            *    70870    70969: gap of unknown length
            *    70970    77012: contig of 6043 bp in length
            *    77013    77112: gap of unknown length
            *    77113    82785: contig of 5673 bp in length
            *    82786    82885: gap of unknown length
            *    82886    88405: contig of 5520 bp in length
            *    88406    88505: gap of unknown length
            *    88506    93767: contig of 5262 bp in length
            *    93768    93867: gap of unknown length
            *    93868    99572: contig of 5705 bp in length
            *    99573    99672: gap of unknown length
            *    99673   109490: contig of 9818 bp in length
            *   109491   109590: gap of unknown length
            *   109591   119557: contig of 9967 bp in length
            *   119558   119657: gap of unknown length
            *   119658   131547: contig of 11890 bp in length
            *   131548   131647: gap of unknown length
            *   131648   148258: contig of 16611 bp in length
            *   148259   148358: gap of unknown length
            *   148359   162560: contig of 14202 bp in length.
FEATURES             Location/Qualifiers
     source          1. .162560
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /clone="RP11-712I15"
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 2;  Length 162560;
  Best Local Similarity   90.5%;  Pred. No. 4.1e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              ||||||||||| | ||||||
Db      64559 CAGTGACATGCAAGACTAGCT 64539
```

```
RESULT 14
AC103719/c
LOCUS       AC103719              167878 bp    DNA     linear   PRI 07-JAN-2003
DEFINITION  Homo sapiens chromosome 8, clone RP11-421P23, complete sequence.
ACCESSION   AC103719
VERSION     AC103719.12  GI:27531859
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 167878)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-421P23
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 167878)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 167878)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JAN-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 167878)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
```

```
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JAN-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jan 7, 2003 this sequence version replaced gi:27476180.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L21648
                Center clone name: 421_P_23
            --------------.
FEATURES             Location/Qualifiers
     source          1. .167878
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-421P23"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   complement(31. .462)
                     /rpt_family="MLT1C"
     repeat_region   complement(894. .996)
                     /rpt_family="L1MEd"
     repeat_region   1036. .1346
                     /rpt_family="AluSx"
     repeat_region   complement(3464. .3610)
                     /rpt_family="L3"
     repeat_region   3612. .3691
                     /rpt_family="MER91A"
     repeat_region   complement(3889. .4086)
                     /rpt_family="L3"
     repeat_region   complement(4429. .4496)
                     /rpt_family="MIR"
     repeat_region   complement(4945. .6217)
                     /rpt_family="HSMAR2"
     repeat_region   complement(6639. .6936)
                     /rpt_family="AluSx"
     repeat_region   7496. .7806
                     /rpt_family="AluJb"
     repeat_region   complement(8574. .8660)
                     /rpt_family="MIR3"
     repeat_region   complement(9589. .9697)
                     /rpt_family="MIR"
     repeat_region   9913. .10223
                     /rpt_family="AluY"
     repeat_region   11320. .11369
                     /rpt_family="(GAAA)n"
     repeat_region   complement(11413. .11593)
                     /rpt_family="L1PA4"
     repeat_region   11803. .11885
                     /rpt_family="(TAGA)n"
     repeat_region   12730. .12830
                     /rpt_family="L2"
     repeat_region   13706. .14005
                     /rpt_family="AluSx"
     repeat_region   15664. .15770
                     /rpt_family="MIR3"
     repeat_region   16008. .16046
                     /rpt_family="L3"
     repeat_region   16183. .16300
                     /rpt_family="MIR"
     repeat_region   complement(19228. .19381)
                     /rpt_family="AluSg/x"
     repeat_region   19858. .21101
                     /rpt_family="L1ME1"
     repeat_region   21178. .21327
                     /rpt_family="L1ME1"
     repeat_region   complement(21343. .21852)
                     /rpt_family="L1P4"
     repeat_region   21853. .22231
                     /rpt_family="L1PA10"
     repeat_region   22235. .22332
                     /rpt_family="L1ME1"
     repeat_region   complement(22344. .22473)
                     /rpt_family="FLAM_C"
     repeat_region   22508. .22555
                     /rpt_family="(TA)n"
     repeat_region   22561. .22636
                     /rpt_family="L1ME2"
     repeat_region   22704. .22828
                     /rpt_family="L2"
     repeat_region   23023. .23203
                     /rpt_family="L2"
     repeat_region   23455. .23508
                     /rpt_family="(TA)n"
     repeat_region   23693. .24079
                     /rpt_family="MLT1I"
     repeat_region   24286. .24330
                     /rpt_family="GA-rich"
     repeat_region   24427. .24468
                     /rpt_family="AT_rich"
     repeat_region   24986. .25013
                     /rpt_family="(TG)n"
     repeat_region   complement(25041. .25110)
                     /rpt_family="MIR"
     repeat_region   25148. .25191
                     /rpt_family="polypurine"
     repeat_region   25198. .25254
                     /rpt_family="CT-rich"
     repeat_region   complement(26391. .26594)
                     /rpt_family="MLT1J2"
     repeat_region   26702. .27267
                     /rpt_family="MLT1E"
     repeat_region   27434. .27468
                     /rpt_family="(TA)n"
     repeat_region   28895. .29360
                     /rpt_family="LTR33"
     repeat_region   29723. .35996
                     /rpt_family="L1PA7"
     repeat_region   complement(36434. .36707)
                     /rpt_family="AluJo"
     repeat_region   37116. .37136
                     /rpt_family="AT_rich"
     repeat_region   complement(38733. .39144)

  Query Match             84.8%;  Score 17.8;  DB 9;  Length 167878;
  Best Local Similarity   90.5%;  Pred. No. 4.1e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              ||||||||||||| | ||||||
Db        842 CAGTGACATGCAAGACTAGCT 822


RESULT 15
AC015468/c
LOCUS       AC015468              189662 bp    DNA     linear   PRI 01-MAY-2001
DEFINITION  Homo sapiens chromosome 8, clone RP11-369E15, complete sequence.
ACCESSION   AC015468
VERSION     AC015468.5  GI:13899433
```

```
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 189662)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-369E15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 189662)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 189662)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
            Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
            Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
            Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 1, 2001 this sequence version replaced gi:12313808.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L2466
                Center clone name: 369_E_15
            --------------.
FEATURES             Location/Qualifiers
     source          1. .189662
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-369E15"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   1. .100
                     /rpt_family="MLT1A1"
     repeat_region   1594. .1741
                     /rpt_family="GA-rich"
     repeat_region   complement(2782. .2957)
                     /rpt_family="MIR"
     repeat_region   3311. .3471
                     /rpt_family="MIR"
     repeat_region   3964. .4067
                     /rpt_family="MIR"
     repeat_region   complement(4528. .4771)
                     /rpt_family="L2"
     repeat_region   4776. .4871
                     /rpt_family="MER4A"
     repeat_region   4869. .5223
                     /rpt_family="MER4A"
     repeat_region   complement(5255. .5765)
                     /rpt_family="L2"
     repeat_region   complement(5805. .6003)
                     /rpt_family="THE1A"
     repeat_region   6004. .6035
                     /rpt_family="(TTA)n"
     repeat_region   complement(6151. .6319)
                     /rpt_family="SVA"
     repeat_region   complement(6320. .6468)
                     /rpt_family="THE1A"
     repeat_region   8722. .8751
                     /rpt_family="(TTTTA)n"
     repeat_region   9155. .9553
                     /rpt_family="MSTB"
     repeat_region   9599. .9982
                     /rpt_family="THE1C"
     repeat_region   complement(10113. .10320)
                     /rpt_family="AluY"
     repeat_region   10822. .10970
                     /rpt_family="MIR3"
     unsure          11210. .11215
                     /note="<30 qual SNGL region"
     repeat_region   11794. .12080
                     /rpt_family="AluSx"
     repeat_region   12974. .13010
                     /rpt_family="(TG)n"
     repeat_region   complement(13091. .13304)
                     /rpt_family="MER106A"
     repeat_region   13484. .13858
                     /rpt_family="MLT1C"
     repeat_region   14025. .14260
                     /rpt_family="L2"
     repeat_region   14584. .14679
                     /rpt_family="MIR"
     repeat_region   14722. .14825
                     /rpt_family="L2"
     repeat_region   15167. .15332
                     /rpt_family="MER33"
     repeat_region   15436. .15788
                     /rpt_family="L2"
     repeat_region   complement(16078. .16165)
                     /rpt_family="AluJ/FRAM"
     repeat_region   16532. .17094
                     /rpt_family="MLT1F1"
     repeat_region   17653. .17674
                     /rpt_family="(TAA)n"
     repeat_region   complement(17675. .17892)
                     /rpt_family="MER7C"
     repeat_region   complement(17897. .18650)
                     /rpt_family="LTR17"
     repeat_region   complement(18651. .18996)
                     /rpt_family="MER7C"
     repeat_region   18999. .19103
                     /rpt_family="(TAA)n"
     repeat_region   complement(19203. .19283)
                     /rpt_family="MIR"
```

```
     repeat_region   20464. .20523
                     /rpt_family="A-rich"
     repeat_region   complement(21279. .21902)
                     /rpt_family="MER41B"
     repeat_region   21922. .22092
                     /rpt_family="MIR"
     repeat_region   22249. .22373
                     /rpt_family="MIR"
     repeat_region   23956. .24058
                     /rpt_family="L2"
     repeat_region   complement(24069. .24198)
                     /rpt_family="MIR"
     repeat_region   25149. .25213
                     /rpt_family="L2"
     repeat_region   complement(25783. .26130)
                     /rpt_family="MLT1I"
     repeat_region   complement(26174. .26539)
                     /rpt_family="MLT1A1"
     repeat_region   complement(28187. .28528)
                     /rpt_family="AluY"
     repeat_region   complement(28792. .29263)
                     /rpt_family="Charlie8"
     repeat_region   complement(30652. .30730)
                     /rpt_family="MER53"
     repeat_region   complement(30986. .31117)
                     /rpt_family="L1ME1"
     repeat_region   31113. .31556
                     /rpt_family="L1ME2"
     repeat_region   31590. .32174
                     /rpt_family="L1ME2"
     repeat_region   32162. .32588
                     /rpt_family="L1ME3"
     repeat_region   33242. .33490
                     /rpt_family="MIR"
     repeat_region   complement(34416. .34803)
                     /rpt_family="MSTB"
     repeat_region   complement(35044. .35350)
                     /rpt_family="AluSg"
     repeat_region   36407. .36695
                     /rpt_family="AluSx"
     repeat_region   complement(37670. .37790)
                     /rpt_family="FLAM_A"
     repeat_region   complement(38065. .38132)
                     /rpt_family="L2"
     repeat_region   complement(38504. .38590)
                     /rpt_family="MIR"
     repeat_region   39044. .39228
                     /rpt_family="MIR3"
     repeat_region   complement(40317. .40473)
                     /rpt_family="MIR"

  Query Match             84.8%;  Score 17.8;  DB 9;  Length 189662;
  Best Local Similarity   90.5%;  Pred. No. 4e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              ||||||||||| | ||||||
Db     128559 CAGTGACATGCAAGACTAGCT 128539


RESULT 16
AC129792
LOCUS       AC129792              197796 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-304B3, *** SEQUENCING IN PROGRESS
            ***, 5 unordered pieces.
ACCESSION   AC129792
VERSION     AC129792.4  GI:25073629
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 197796)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 197796)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 197796)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 19, 2002 this sequence version replaced gi:23915295.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
```

```
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: KATF
                Center clone name: CH230-304B3
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 191737 bases at least Q40
                Consensus quality: 193479 bases at least Q30
                Consensus quality: 194485 bases at least Q20
                Estimated insert size: 194733; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   191071: contig of 191071 bp in length
            *  191072   191171: gap of unknown length
            *  191172   192706: contig of 1535 bp in length
            *  192707   192806: gap of unknown length
            *  192807   194126: contig of 1320 bp in length
            *  194127   194226: gap of unknown length
            *  194227   196128: contig of 1902 bp in length
            *  196129   196228: gap of unknown length
            *  196229   197796: contig of 1568 bp in length.
FEATURES             Location/Qualifiers
     source          1. .197796
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-304B3"
     misc_feature    1. .1063
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    1565. .2730
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    3873. .4716
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:BZ204887"
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 2;  Length 197796;
  Best Local Similarity   90.5%;  Pred. No. 4e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              | |||| ||||||||||||||
Db     192834 CTGTGATATGCAGGTCTAGCT 192854


RESULT 17
AC129792/c
LOCUS       AC129792               197796 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-304B3, *** SEQUENCING IN PROGRESS
            ***, 5 unordered pieces.
ACCESSION   AC129792
VERSION     AC129792.4  GI:25073629
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 197796)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 197796)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 197796)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 19, 2002 this sequence version replaced gi:23915295.
            The sequence in this assembly is a combination of BAC based reads
```

```
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: KATF
                Center clone name: CH230-304B3
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 191737 bases at least Q40
                Consensus quality: 193479 bases at least Q30
                Consensus quality: 194485 bases at least Q20
                Estimated insert size: 194733; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   191071: contig of 191071 bp in length
            *  191072   191171: gap of unknown length
            *  191172   192706: contig of 1535 bp in length
            *  192707   192806: gap of unknown length
            *  192807   194126: contig of 1320 bp in length
            *  194127   194226: gap of unknown length
            *  194227   196128: contig of 1902 bp in length
            *  196129   196228: gap of unknown length
            *  196229   197796: contig of 1568 bp in length.
FEATURES             Location/Qualifiers
     source          1. .197796
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-304B3"
     misc_feature    1. .1063
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    1565. .2730
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    3873. .4716
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:BZ204887"
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 2;  Length 197796;
  Best Local Similarity   90.5%;  Pred. No. 4e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              | |||| ||||||||||||||
Db      97483 CTGTGATATGCAGGTCTAGCT 97463
```

```
RESULT 18
AC087221
LOCUS       AC087221              203690 bp    DNA     linear   HTG 24-MAY-2002
DEFINITION  Homo sapiens chromosome 8 clone RP11-712I15 map 8, WORKING DRAFT
            SEQUENCE, 34 ordered pieces.
ACCESSION   AC087221
VERSION     AC087221.2  GI:21166223
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 203690)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-712I15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 203690)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-DEC-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 203690)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
            Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
            Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
            Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAY-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 24, 2002 this sequence version replaced gi:11875303.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
```

```
    Center: Whitehead Institute/ MIT Center for Genome Research
    Center code: WIBR
    Web site: http://www-seq.wi.mit.edu
    Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
    Center project name: L11638
    Center clone name: 712_I_15
-------------- Summary Statistics
    Sequencing vector: Plasmid; n/a; 100% of reads
    Chemistry: Dye-terminator Big Dye; 100% of reads
    Assembly program: Phrap; version 0.960731
    Consensus quality: 190084 bases at least Q40
    Consensus quality: 196705 bases at least Q30
    Consensus quality: 199369 bases at least Q20
    Insert size: 176000; agarose-fp
    Insert size: 200390; sum-of-contigs
    Quality coverage: 9.6 in Q20 bases; agarose-fp
    Quality coverage: 8.5 in Q20 bases; sum-of-contigs
--------------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 34 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*        1      460: contig of 460 bp in length
*      461      560: gap of 100 bp
*      561     1196: contig of 636 bp in length
*     1197     1296: gap of 100 bp
*     1297     2122: contig of 826 bp in length
*     2123     2222: gap of 100 bp
*     2223     2869: contig of 647 bp in length
*     2870     2969: gap of 100 bp
*     2970     3620: contig of 651 bp in length
*     3621     3720: gap of 100 bp
*     3721     4314: contig of 594 bp in length
*     4315     4414: gap of 100 bp
*     4415     5152: contig of 738 bp in length
*     5153     5252: gap of 100 bp
*     5253     6209: contig of 957 bp in length
*     6210     6309: gap of 100 bp
*     6310     7103: contig of 794 bp in length
*     7104     7203: gap of 100 bp
*     7204     7937: contig of 734 bp in length
*     7938     8037: gap of 100 bp
*     8038     9006: contig of 969 bp in length
*     9007     9106: gap of 100 bp
*     9107     9906: contig of 800 bp in length
*     9907    10006: gap of 100 bp
*    10007    11070: contig of 1064 bp in length
*    11071    11170: gap of 100 bp
*    11171    12297: contig of 1127 bp in length
*    12298    12397: gap of 100 bp
*    12398    13415: contig of 1018 bp in length
*    13416    13515: gap of 100 bp
*    13516    14231: contig of 716 bp in length
*    14232    14331: gap of 100 bp
*    14332    15596: contig of 1265 bp in length
*    15597    15696: gap of 100 bp
*    15697    16462: contig of 766 bp in length
*    16463    16562: gap of 100 bp
*    16563    17742: contig of 1180 bp in length
*    17743    17842: gap of 100 bp
*    17843    18857: contig of 1015 bp in length
*    18858    18957: gap of 100 bp
*    18958    20490: contig of 1533 bp in length
*    20491    20590: gap of 100 bp
*    20591    22210: contig of 1620 bp in length
*    22211    22310: gap of 100 bp
*    22311    23851: contig of 1541 bp in length
*    23852    23951: gap of 100 bp
*    23952    25684: contig of 1733 bp in length
*    25685    25784: gap of 100 bp
*    25785    27006: contig of 1222 bp in length
*    27007    27106: gap of 100 bp
*    27107    28616: contig of 1510 bp in length
*    28617    28716: gap of 100 bp
*    28717    30184: contig of 1468 bp in length
*    30185    30284: gap of 100 bp
*    30285    31303: contig of 1019 bp in length
*    31304    31403: gap of 100 bp
*    31404    32917: contig of 1514 bp in length
*    32918    33017: gap of 100 bp
*    33018    42030: contig of 9013 bp in length
*    42031    42130: gap of 100 bp
*    42131    49012: contig of 6882 bp in length
*    49013    49112: gap of 100 bp
*    49113    62240: contig of 13128 bp in length
*    62241    62340: gap of 100 bp
*    62341    86795: contig of 24455 bp in length
*    86796    86895: gap of 100 bp
*    86896   203690: contig of 116795 bp in length.
FEATURES             Location/Qualifiers
     source          1. .203690
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-712I15"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .460
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    561. .1196
                     /note="assembly_fragment"
     misc_feature    1297. .2122
                     /note="assembly_fragment"
     misc_feature    2223. .2869
                     /note="assembly_fragment"
     misc_feature    2970. .3620
                     /note="assembly_fragment"
     misc_feature    3721. .4314
                     /note="assembly_fragment"
     misc_feature    4415. .5152
                     /note="assembly_fragment"
     misc_feature    5253. .6209
                     /note="assembly_fragment"
     misc_feature    6310. .7103
                     /note="assembly_fragment"
     misc_feature    7204. .7937
                     /note="assembly_fragment"
     misc_feature    8038. .9006
                     /note="assembly_fragment"
     misc_feature    9107. .9906
                     /note="assembly_fragment"

  Query Match             84.8%;  Score 17.8;  DB 2;  Length 203690;
  Best Local Similarity   90.5%;  Pred. No. 4e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              ||||||||||||  | ||||||
Db     190273 CAGTGACATGCAAGACTAGCT 190293


RESULT 19
AC115746
LOCUS       AC115746               214765 bp    DNA     linear   ROD 29-JUL-2004
DEFINITION  Mus musculus chromosome 15, clone RP23-3J8, complete sequence.
ACCESSION   AC115746
VERSION     AC115746.10  GI:50811761
```

```
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 214765)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 15, clone RP23-3J8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 214765)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 214765)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
            Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J.,
            Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
            DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
            Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,
            Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
            Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
            MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
            Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
            Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUN-2004) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 214765)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
            Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J.,
            Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
            DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
            Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,
            Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
            Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
            MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
            Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
            Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-2004) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 29, 2004 this sequence version replaced gi:48058970.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@broad.mit.edu
            -------------- Project Information
                Center project name: L22826
                Center clone name: 3_J_8
            --------------

            Some of the sequence contained within base pairs 88151 to the end
            of the clone was stolen from accession AC110379.
FEATURES             Location/Qualifiers
     source          1. .214765
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="15"
                     /map="15"
                     /clone="RP23-3J8"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .6
                     /note="clone_boundary
                     clone_end:SP6
                     site:EcoRI"
     repeat_region   645. .709
                     /rpt_family="B4A"
     repeat_region   710. .782
                     /rpt_family="B1F"
     repeat_region   complement(988. .1089)
                     /rpt_family="MIR"
     repeat_region   1234. .1627
                     /rpt_family="RMER10B"
     repeat_region   1696. .1871
                     /rpt_family="B3"
     repeat_region   1906. .2169
                     /rpt_family="B4A"
     repeat_region   3739. .3906
                     /rpt_family="RSINE1"
     repeat_region   5660. .5784
                     /rpt_family="ID_B1"
     repeat_region   6541. .6571
                     /rpt_family="GC_rich"
     repeat_region   complement(6923. .7069)
                     /rpt_family="B1_MM"
     repeat_region   7673. .7749
                     /rpt_family="ID_B1"
     repeat_region   10099. .10216
                     /rpt_family="L1"
     repeat_region   10217. .10354
                     /rpt_family="B1_MM"
     repeat_region   10355. .11140
                     /rpt_family="L1"
     repeat_region   12096. .12140
                     /rpt_family="AT_rich"
     repeat_region   12141. .12257
                     /rpt_family="PB1D9"
```

```
     repeat_region   12370. .12444
                     /rpt_family="(CA)n"
     repeat_region   13343. .13367
                     /rpt_family="(TGGGGG)n"
     repeat_region   14362. .14384
                     /rpt_family="AT_rich"
     repeat_region   complement(15620. .15776)
                     /rpt_family="B1_MM"
     repeat_region   15792. .15822
                     /rpt_family="(TTTC)n"
     repeat_region   complement(15830. .15977)
                     /rpt_family="B1_MM"
     repeat_region   complement(16486. .16564)
                     /rpt_family="ID1_MM"
     repeat_region   complement(17181. .17277)
                     /rpt_family="RSINE1"
     repeat_region   17293. .17388
                     /rpt_family="B4"
     repeat_region   17391. .17584
                     /rpt_family="B2_Mm2"
     repeat_region   17585. .17621
                     /rpt_family="polypurine"
     repeat_region   17623. .17960
                     /rpt_family="(GGA)n"
     repeat_region   18167. .18196
                     /rpt_family="(CAAA)n"
     repeat_region   18533. .18595
                     /rpt_family="ID5"
     repeat_region   18596. .18608
                     /rpt_family="ID_B1"
     repeat_region   19265. .19413
                     /rpt_family="B1_MM"
     repeat_region   19715. .19921
                     /rpt_family="B4A"
     repeat_region   19952. .19992
                     /rpt_family="(CA)n"
     repeat_region   20016. .20141
                     /rpt_family="PB1D9"
     repeat_region   20286. .20398
                     /rpt_family="B1_MM"
     repeat_region   20399. .20488
                     /rpt_family="GA-rich"
     repeat_region   20736. .20898
                     /rpt_family="B3A"
     repeat_region   20931. .20957
                     /rpt_family="(CAAA)n"
     repeat_region   22176. .22256
                     /rpt_family="(GA)n"
     repeat_region   23275. .23413
                     /rpt_family="B1_MM"

  Query Match             84.8%;  Score 17.8;  DB 10;  Length 214765;
  Best Local Similarity   90.5%;  Pred. No. 3.9e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              ||||||| |||||| ||||||
Db      98096 CAGTGACTTGCAGGGCTAGCT 98116


RESULT 20
AC120123/c
LOCUS       AC120123               222540 bp    DNA     linear   HTG 16-SEP-2004
DEFINITION  Mus musculus chromosome 7 clone RP23-152B12 map 7, *** SEQUENCING
            IN PROGRESS ***, 7 unordered pieces.
ACCESSION   AC120123
VERSION     AC120123.10  GI:52138883
KEYWORDS    HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 222540)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 7, clone RP23-152B12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 222540)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAY-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 222540)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
            Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J.,
            Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
            DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
            Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,
            Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
            Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
            MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
            Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
            Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-SEP-2004) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Sep 16, 2004 this sequence version replaced gi:50284650.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@broad.mit.edu
            -------------- Project Information
                Center project name: L15655
                Center clone name: 152_B_12
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
```

```
            * as soon as it is available and the accession number will
            * be preserved.
            *         1     47731: contig of 47731 bp in length
            *     47732     47831: gap of unknown length
            *     47832     61722: contig of 13891 bp in length
            *     61723     61822: gap of unknown length
            *     61823    153130: contig of 91308 bp in length
            *    153131    153230: gap of unknown length
            *    153231    170552: contig of 17322 bp in length
            *    170553    170652: gap of unknown length
            *    170653    190827: contig of 20175 bp in length
            *    190828    190927: gap of unknown length
            *    190928    216618: contig of 25691 bp in length
            *    216619    216718: gap of unknown length
            *    216719    222540: contig of 5822 bp in length.
FEATURES             Location/Qualifiers
     source          1. .222540
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="7"
                     /map="7"
                     /clone="RP23-152B12"
                     /clone_lib="RPCI-23 Female Mouse BAC"
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 2;  Length 222540;
  Best Local Similarity   90.5%;  Pred. No. 3.9e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              |||||| |||||||| |||||
Db     177745 CAGTGATATGCAGGTGTAGCT 177725


RESULT 21
AC115307
LOCUS       AC115307               260600 bp    DNA     linear   HTG 09-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-11F18, WORKING DRAFT SEQUENCE.
ACCESSION   AC115307
VERSION     AC115307.4  GI:24817861
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 260600)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 260600)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 260600)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 9, 2002 this sequence version replaced gi:22772493.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GEBE
                Center clone name: CH230-11F18
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 229014 bases at least Q40
                Consensus quality: 231787 bases at least Q30
                Consensus quality: 232904 bases at least Q20
                Estimated insert size: 237320; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
```

```
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   260600: contig of 260600 bp in length.
FEATURES             Location/Qualifiers
     source          1. .260600
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-11F18"
     misc_feature    1. .2114
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    6641. .7491
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH340447"
     misc_feature    88160. .195463
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH340449"
     misc_feature    258937. .260600
                     /note="wgs_end_extension
                     clone_end:Sp6"
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 2;  Length 260600;
  Best Local Similarity   90.5%;  Pred. No. 3.9e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              ||||||||||| |||||| |||
Db     178018 CAGTGACATGTAGGTCTTGCT 178038


RESULT 22
AC129380/c
LOCUS       AC129380               281447 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-233D15, *** SEQUENCING IN PROGRESS
            ***.
ACCESSION   AC129380
VERSION     AC129380.3  GI:30578573
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 281447)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S.; Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 281447)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 281447)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:23116981.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GJZM
                Center clone name: CH230-233D15
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 224946 bases at least Q40
                Consensus quality: 227863 bases at least Q30
                Consensus quality: 229407 bases at least Q20
                Estimated insert size: 234544; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
```

```
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   281447: contig of 281447 bp in length.
FEATURES             Location/Qualifiers
     source          1. .281447
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-233D15"
     misc_feature    1. .1083
                     /note="wgs_contig"
     misc_feature    2987. .4464
                     /note="wgs_contig"
     misc_feature    192314. .192414
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BZ091967"
     misc_feature    280368. .281447
                     /note="wgs_end_extension
                     clone_end:Sp6"
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 2;  Length 281447;
  Best Local Similarity   90.5%;  Pred. No. 3.8e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              | |||| ||||||||||||||
Db     264359 CTGTGATATGCAGGTCTAGCT 264339


RESULT 23
BD185177/c
LOCUS       BD185177                5446 bp    DNA     linear   PAT 17-JUN-2003
DEFINITION  Novel genes and proteins encoded by the genes.
ACCESSION   BD185177
VERSION     BD185177.1  GI:31877377
KEYWORDS    JP 2002345493-A/20.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5446)
  AUTHORS   Ohara,O., Nagase,T. and Nakajima,D.
  TITLE     Novel genes and proteins encoded by the genes
  JOURNAL   Patent: JP 2002345493-A 20 03-DEC-2002;
            KAZUSA DNA RESEARCH INSTITUTE
COMMENT     OS   Homo sapiens (human)
            PN   JP 2002345493-A/20
            PD   03-DEC-2002
            PF   26-FEB-2002 JP 2002049046
            PI   OSAMU OHARA,TAKAHIRO NAGASE,DAISUKE NAKAJIMA
            PC   C12N15/09,C07K14/47,C07K14/54,C12N15/00
            CC   Novel genes and proteins encoded by the genes FH   Key
                 Location/Qualifiers
            FT   CDS             (2564). .(4138).
FEATURES             Location/Qualifiers
     source          1. .5446
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             82.9%;  Score 17.4;  DB 6;  Length 5446;
  Best Local Similarity   94.7%;  Pred. No. 9.5e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAG 19
              |||||||| ||||||||||
Db        833 CAGTGACAGGCAGGTCTAG 815


RESULT 24
AC121561/c
LOCUS       AC121561               46070 bp    DNA     linear   HTG 20-MAY-2002
DEFINITION  Homo sapiens chromosome 17 clone CTD-2527E13 map 17, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC121561
VERSION     AC121561.1  GI:20986629
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 46070)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone CTD-2527E13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 46070)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
            Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
            Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
            Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAY-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L26773
                Center clone name: 2527_E_13
            --------------.
            * NOTE: This record contains 58 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
```

```
*       1      684: contig of 684 bp in length
*     685      784: gap of 100 bp
*     785     1401: contig of 617 bp in length
*    1402     1501: gap of 100 bp
*    1502     2203: contig of 702 bp in length
*    2204     2303: gap of 100 bp
*    2304     2984: contig of 681 bp in length
*    2985     3084: gap of 100 bp
*    3085     3754: contig of 670 bp in length
*    3755     3854: gap of 100 bp
*    3855     4545: contig of 691 bp in length
*    4546     4645: gap of 100 bp
*    4646     5346: contig of 701 bp in length
*    5347     5446: gap of 100 bp
*    5447     6142: contig of 696 bp in length
*    6143     6242: gap of 100 bp
*    6243     6949: contig of 707 bp in length
*    6950     7049: gap of 100 bp
*    7050     7730: contig of 681 bp in length
*    7731     7830: gap of 100 bp
*    7831     8515: contig of 685 bp in length
*    8516     8615: gap of 100 bp
*    8616     9267: contig of 652 bp in length
*    9268     9367: gap of 100 bp
*    9368    10068: contig of 701 bp in length
*   10069    10168: gap of 100 bp
*   10169    10859: contig of 691 bp in length
*   10860    10959: gap of 100 bp
*   10960    11667: contig of 708 bp in length
*   11668    11767: gap of 100 bp
*   11768    12472: contig of 705 bp in length
*   12473    12572: gap of 100 bp
*   12573    13277: contig of 705 bp in length
*   13278    13377: gap of 100 bp
*   13378    14069: contig of 692 bp in length
*   14070    14169: gap of 100 bp
*   14170    14870: contig of 701 bp in length
*   14871    14970: gap of 100 bp
*   14971    15681: contig of 711 bp in length
*   15682    15781: gap of 100 bp
*   15782    16482: contig of 701 bp in length
*   16483    16582: gap of 100 bp
*   16583    17271: contig of 689 bp in length
*   17272    17371: gap of 100 bp
*   17372    18068: contig of 697 bp in length
*   18069    18168: gap of 100 bp
*   18169    18863: contig of 695 bp in length
*   18864    18963: gap of 100 bp
*   18964    19650: contig of 687 bp in length
*   19651    19750: gap of 100 bp
*   19751    20456: contig of 706 bp in length
*   20457    20556: gap of 100 bp
*   20557    21267: contig of 711 bp in length
*   21268    21367: gap of 100 bp
*   21368    22078: contig of 711 bp in length
*   22079    22178: gap of 100 bp
*   22179    22875: contig of 697 bp in length
*   22876    22975: gap of 100 bp
*   22976    23661: contig of 686 bp in length
*   23662    23761: gap of 100 bp
*   23762    24488: contig of 727 bp in length
*   24489    24588: gap of 100 bp
*   24589    25276: contig of 688 bp in length
*   25277    25376: gap of 100 bp
*   25377    26082: contig of 706 bp in length
*   26083    26182: gap of 100 bp
*   26183    26886: contig of 704 bp in length
*   26887    26986: gap of 100 bp
*   26987    27691: contig of 705 bp in length
*   27692    27791: gap of 100 bp
*   27792    28494: contig of 703 bp in length
*   28495    28594: gap of 100 bp
*   28595    29303: contig of 709 bp in length
*   29304    29403: gap of 100 bp
*   29404    30080: contig of 677 bp in length
*   30081    30180: gap of 100 bp
*   30181    30844: contig of 664 bp in length
*   30845    30944: gap of 100 bp
*   30945    31640: contig of 696 bp in length
*   31641    31740: gap of 100 bp
*   31741    32459: contig of 719 bp in length
*   32460    32559: gap of 100 bp
*   32560    33262: contig of 703 bp in length
*   33263    33362: gap of 100 bp
*   33363    34090: contig of 728 bp in length
*   34091    34190: gap of 100 bp
*   34191    34892: contig of 702 bp in length
*   34893    34992: gap of 100 bp
*   34993    35692: contig of 700 bp in length
*   35693    35792: gap of 100 bp
*   35793    36480: contig of 688 bp in length
*   36481    36580: gap of 100 bp
*   36581    37293: contig of 713 bp in length
*   37294    37393: gap of 100 bp
*   37394    38091: contig of 698 bp in length
*   38092    38191: gap of 100 bp
*   38192    38897: contig of 706 bp in length
*   38898    38997: gap of 100 bp
*   38998    39720: contig of 723 bp in length
*   39721    39820: gap of 100 bp
*   39821    40526: contig of 706 bp in length
*   40527    40626: gap of 100 bp
*   40627    41302: contig of 676 bp in length
*   41303    41402: gap of 100 bp
*   41403    42077: contig of 675 bp in length
*   42078    42177: gap of 100 bp
*   42178    42841: contig of 664 bp in length
*   42842    42941: gap of 100 bp
*   42942    43638: contig of 697 bp in length
*   43639    43738: gap of 100 bp
*   43739    44456: contig of 718 bp in length
*   44457    44556: gap of 100 bp
*   44557    45255: contig of 699 bp in length
*   45256    45355: gap of 100 bp
*   45356    46070: contig of 715 bp in length.
FEATURES             Location/Qualifiers
     source          1. .46070
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="CTD-2527E13"
                     /clone_lib="CITD2 Human BAC"
ORIGIN

  Query Match             82.9%;  Score 17.4;  DB 2;  Length 46070;
  Best Local Similarity   94.7%;  Pred. No. 7.4e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAG 19
              |||||||| ||||||||||
Db      31187 CAGTGACAGGCAGGTCTAG 31169


RESULT 25
AL133227
LOCUS       AL133227               85566 bp    DNA     linear   PRI 04-APR-2001
DEFINITION  Human DNA sequence from clone RP11-394O2 on chromosome 20. Contains
            the gene for CGI-15 protein, a gene for a novel protein similar to
            KIAA0281 and Drosophila CG5336, ESTs, STSs, GSSs and a CpG island,
            complete sequence.
ACCESSION   AL133227
VERSION     AL133227.15  GI:9187135
KEYWORDS    HTG; CpG island; KIAA0281.
SOURCE      Homo sapiens (human)
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 85566)
  AUTHORS   Barlow,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 14, 2000 this sequence version replaced gi:8977995.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 20, constructed by the Sanger Centre Chromosome 20
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr20
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-394O2 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP11-394O2 is at 1 in this sequence. The
            true left end of clone RP5-981L23 is at 85467 in this sequence. The
            true right end of clone RP5-984I23 is at 78450 in this sequence.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. RP11-394O2 is from
            the library RPCI-11.2 constructed by the group of Pieter de Jong.
            For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .85566
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /clone="RP11-394O2"
                     /clone_lib="RPCI-11.2"
     repeat_region   33. .90
                     /note="MIR repeat: matches 107. .161 of consensus"
     repeat_region   139. .438
                     /note="AluSx repeat: matches 1. .293 of consensus"
     repeat_region   1404. .1535
                     /note="MIR repeat: matches 54. .182 of consensus"
     repeat_region   3358. .3536
                     /note="L2 repeat: matches 2318. .2500 of consensus"
     misc_feature    3438. .3969
                     /note="match: GSS: Em:AQ373404"
     misc_feature    3456. .3976
                     /note="match: GSS: Em:AQ885421"
     repeat_region   3620. .3905
                     /note="MIR repeat: matches 16. .262 of consensus"
     misc_feature    3653. .4183
                     /note="match: GSS: Em:AQ374823"
     repeat_region   5271. .5327
                     /note="MIR repeat: matches 61. .121 of consensus"
     repeat_region   6122. .6156
                     /note="MIR repeat: matches 110. .144 of consensus"
     repeat_region   6577. .6685
                     /note="L1M4 repeat: matches 5209. .5334 of consensus"
     repeat_region   7049. .7178
                     /note="L2 repeat: matches 2562. .2695 of consensus"
     repeat_region   7275. .7472
                     /note="MIR repeat: matches 44. .254 of consensus"
     repeat_region   7545. .7917
                     /note="LTR16C repeat: matches 11. .381 of consensus"
     misc_feature    7733. .8207
                     /note="match: GSS: Em:AQ524007"
     repeat_region   8425. .8510
                     /note="MIR repeat: matches 39. .148 of consensus"
     repeat_region   8622. .8835
                     /note="MIR repeat: matches 35. .262 of consensus"
     repeat_region   8945. .9114
                     /note="MIR repeat: matches 21. .210 of consensus"
     repeat_region   9536. .9599
                     /note="MIR repeat: matches 118. .181 of consensus"
     repeat_region   9666. .9851
                     /note="MIR repeat: matches 80. .262 of consensus"
     misc_feature    10546. .10783
                     /note="match: GSS: Em:AQ543754"
     repeat_region   11352. .11438
                     /note="MIR repeat: matches 85. .168 of consensus"
     misc_feature    12303. .12620
                     /note="match: GSS: Em:AQ195196
                     match: STS: Em:G52843"
     misc_feature    12324. .12752
                     /note="match: GSS: Em:AQ631329"
     repeat_region   12855. .13015
                     /note="MIR repeat: matches 65. .220 of consensus"
     repeat_region   13723. .14037
                     /note="AluJb repeat: matches 16. .312 of consensus"
     misc_feature    complement(13988. .14650)
                     /note="match: GSS: Em:AQ899717"
     misc_feature    complement(14441. .14659)
                     /note="match: GSS: Em:AQ062441"
     repeat_region   15533. .15742
                     /note="MIR repeat: matches 62. .256 of consensus"
     repeat_region   16607. .16778
                     /note="MER91A repeat: matches 2. .171 of consensus"
     repeat_region   17496. .17568
                     /note="L2 repeat: matches 2624. .2696 of consensus"
     repeat_region   17746. .18049
                     /note="AluSx repeat: matches 1. .304 of consensus"
     repeat_region   18067. .18413
                     /note="L1PA13 repeat: matches 5807. .6156 of consensus"
     repeat_region   18593. .18825
                     /note="MIR repeat: matches 20. .248 of consensus"
     repeat_region   18982. .19111
                     /note="MIR repeat: matches 136. .250 of consensus"
     repeat_region   19112. .19406
                     /note="AluJb repeat: matches 1. .296 of consensus"
     repeat_region   19407. .19518
                     /note="MIR repeat: matches 20. .136 of consensus"
     repeat_region   20100. .20283
                     /note="L2 repeat: matches 2550. .2750 of consensus"
     repeat_region   20312. .20763
                     /note="Charlie4a repeat: matches 17. .485 of consensus"
     repeat_region   20787. .21340
                     /note="L2 repeat: matches 1936. .2534 of consensus"
     repeat_region   21447. .21736
                     /note="MLT1J repeat: matches 1. .320 of consensus"
     repeat_region   21990. .22289
                     /note="AluY repeat: matches 1. .301 of consensus"
     repeat_region   22377. .22462
                     /note="L2 repeat: matches 2621. .2710 of consensus"
     repeat_region   22463. .22749
                     /note="AluSx repeat: matches 1. .292 of consensus"
     repeat_region   24255. .24319
                     /note="MIR repeat: matches 109. .175 of consensus"
     repeat_region   24669. .24768
                     /note="MIR repeat: matches 80. .183 of consensus"
     repeat_region   24832. .24928
                     /note="MIR repeat: matches 92. .190 of consensus"
     repeat_region   26048. .26173
```

```
                     /note="MIR repeat: matches 79. .206 of consensus"
     repeat_region   26436. .26626
                     /note="L2 repeat: matches 2340. .2529 of consensus"
     repeat_region   26684. .26740
                     /note="3 copies 19 mer 91% conserved"
     repeat_region   26722. .26765
                     /note="11 copies 4 mer catc 93% conserved"
     repeat_region   26766. .26809
                     /note="11 copies 4 mer atcc 81% conserved"
     repeat_region   26813. .26844
                     /note="8 copies 4 mer atcc 93% conserved"
     repeat_region   26839. .26918
                     /note="4 copies 20 mer 77% conserved"
     repeat_region   27222. .27559
                     /note="LTR16A repeat: matches 95. .440 of consensus"
     repeat_region   27583. .27748
                     /note="FRAM repeat: matches 1. .166 of consensus"
     repeat_region   27786. .27897
                     /note="L2 repeat: matches 2637. .2750 of consensus"
     repeat_region   28308. .28431
                     /note="AluSq repeat: matches 1. .132 of consensus"
     repeat_region   28448. .28524
                     /note="LTR16A repeat: matches 193. .270 of consensus"
     repeat_region   28525. .28714
                     /note="AluSq repeat: matches 114. .301 of consensus"
     repeat_region   28837. .29017
                     /note="MIR repeat: matches 65. .250 of consensus"
     repeat_region   29053. .29187
                     /note="MIR repeat: matches 84. .226 of consensus"
     misc_feature    30560. .31063
                     /note="match: GSS: Em:AQ172242"
     misc_feature    30596. .30931
                     /note="match: GSS: Em:AQ901120"
     repeat_region   30602. .30764
                     /note="MER58A repeat: matches 46. .221 of consensus"
     gene            complement(31347. .46204)
                     /gene="bA394O2.1"
     mRNA            complement(join(31347. .32333,32570. .32699,33860. .34035,
                     36691. .36774,36882. .37050,37611. .37683,38383. .38446,
                     39432. .39582,40196. .40572,46180. .46204))
                     /gene="bA394O2.1"
                     /product="bA394O2.1 (CGI-15 protein)"
                     /note="match: cDNAs: Em:AF132949
                     match: ESTs: Em:AU079917 Em:AA667893 Em:AI007286
                     Em:AU067617 Em:W56183 Em:R69763 Em:H06603 Em:AW631237

  Query Match             82.9%;  Score 17.4;  DB 9;  Length 85566;
  Best Local Similarity   94.7%;  Pred. No. 6.9e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAG 19
              ||||||||  ||||||||||
Db      64014 CAGTGACAGGCAGGTCTAG 64032


Search completed: September  6, 2005, 20:29:56
Job time : 746.656 secs
```

OM nucleic - nucleic search, using sw model

Run on: September 6, 2005, 17:45:55 ; Search time 1500.84 Seconds
(without alignments)
532.600 Million cell updates/sec

Title: US-10-729-421-40
Perfect score: 21
Sequence: 1 cagtgacatgcaggtctagct 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	19	90.5	1041	9	CNS04CAJ	AL284212 Tetraodon
	2	18.4	87.6	1157	6	CD500170	CD500170 CDA43-E06
	3	17.8	84.8	211	8	AZ995872	AZ995872 2M0281B23
c	4	17.8	84.8	611	8	AZ218630	AZ218630 Sheared D
	5	17.8	84.8	878	9	CR794665	CR794665 GR0AAA12A
c	6	17.8	84.8	4261	3	AK083880	AK083880 Mus muscu
c	7	17.4	82.9	380	9	CL289600	CL289600 ZMMBBb064
	8	17.4	82.9	718	9	CG176005	CG176005 PUIMY95TD
	9	17.4	82.9	722	2	BF607360	BF607360 MY1_00030
c	10	17.4	82.9	772	8	CC340825	CC340825 OGQAQ84TH
	11	17.4	82.9	772	9	CG211386	CG211386 OG5CS03TC
c	12	17.4	82.9	805	9	CG362883	CG362883 OGYBS24TH
c	13	17.4	82.9	816	4	BI548151	BI548151 603189492
	14	17.4	82.9	817	8	CC340835	CC340835 OGQAQ84TV
c	15	17.4	82.9	923	9	CG211861	CG211861 OG1AW29TV
	16	17	81.0	382	6	BY609340	BY609340 BY609340
	17	17	81.0	408	6	BY649695	BY649695 BY649695
c	18	17	81.0	622	6	CA124343	CA124343 SCQGLR108
	19	17	81.0	633	8	AZ069535	AZ069535 RPCI-23-4
c	20	17	81.0	642	2	BB650662	BB650662 BB650662
c	21	17	81.0	655	2	BB293162	BB293162 BB293162
	22	17	81.0	768	9	AG457137	AG457137 Mus muscu
	23	17	81.0	781	9	AG557182	AG557182 Mus muscu
	24	17	81.0	949	7	CF411086	CF411086 CH3#071_D
c	25	17	81.0	2445	3	AK044974	AK044974 Mus muscu
c	26	17	81.0	2466	3	AK013040	AK013040 Mus muscu
c	27	17	81.0	2559	3	AK017012	AK017012 Mus muscu
c	28	17	81.0	3256	3	AK082079	AK082079 Mus muscu
c	29	17	81.0	3417	3	AK081942	AK081942 Mus muscu
c	30	16.8	80.0	306	2	AW846933	AW846933 RC3-CT019
	31	16.8	80.0	330	1	AU249500	AU249500 AU249500
	32	16.8	80.0	335	1	AA050319	AA050319 mj14a05.r
	33	16.8	80.0	387	5	BY077663	BY077663 BY077663
	34	16.8	80.0	400	1	AA004028	AA004028 mg80g02.r
	35	16.8	80.0	400	1	AA117056	AA117056 mn29c06.r
	36	16.8	80.0	407	1	AA003782	AA003782 mg62e10.r
	37	16.8	80.0	407	2	BE656207	BE656207 UI-M-BH0-
	38	16.8	80.0	414	1	AA052553	AA052553 mc66d04.r
	39	16.8	80.0	419	7	W79973	W79973 me90e09.r1
	40	16.8	80.0	426	1	AA008849	AA008849 mg98f08.r
	41	16.8	80.0	435	5	BY273688	BY273688 BY273688
	42	16.8	80.0	436	5	BY240303	BY240303 BY240303
	43	16.8	80.0	442	5	BY051479	BY051479 BY051479
	44	16.8	80.0	443	6	CB788954	CB788954 AMGNNUC:M
	45	16.8	80.0	461	5	BX529253	BX529253 BX529253
	46	16.8	80.0	466	7	W89580	W89580 mf73f08.r1
	47	16.8	80.0	467	1	AJ647590	AJ647590 AJ647590
	48	16.8	80.0	467	2	BB863212	BB863212 BB863212
	49	16.8	80.0	469	1	AJ647892	AJ647892 AJ647892
	50	16.8	80.0	508	6	CA533118	CA533118 C0345E08-
	51	16.8	80.0	510	7	W89380	W89380 mf73g08.r1
	52	16.8	80.0	512	2	BF556581	BF556581 UI-R-E1-f
	53	16.8	80.0	515	2	BE374887	BE374887 601226811
c	54	16.8	80.0	523	8	BZ132704	BZ132704 CH230-481
	55	16.8	80.0	542	1	AA712019	AA712019 vu60d06.r
	56	16.8	80.0	544	7	CF732719	CF732719 UI-M-HA0-
c	57	16.8	80.0	546	8	BH347218	BH347218 CH230-42D
	58	16.8	80.0	549	4	BG276278	BG276278 uv02e10.y
	59	16.8	80.0	552	1	AA682125	AA682125 vu13b03.r
	60	16.8	80.0	561	1	AA530593	AA530593 vj49g09.r
	61	16.8	80.0	568	2	BF452552	BF452552 maa01d05.
	62	16.8	80.0	570	1	AV597290	AV597290 AV597290
	63	16.8	80.0	576	1	AV595391	AV595391 AV595391
	64	16.8	80.0	582	1	AL792667	AL792667 AL792667
	65	16.8	80.0	584	5	BQ840540	BQ840540 mah69f08.
	66	16.8	80.0	588	9	CG672097	CG672097 RRM266 Ba
	67	16.8	80.0	596	4	BG100744	BG100744 uy14c01.y
	68	16.8	80.0	600	4	BI985696	BI985696 3144-63 M
	69	16.8	80.0	617	2	AW412020	AW412020 uo55h02.y
	70	16.8	80.0	619	7	CR421241	CR421241 CR421241
	71	16.8	80.0	624	7	CK621976	CK621976 ml31a12.y
	72	16.8	80.0	626	4	BJ774417	BJ774417 BJ774417
	73	16.8	80.0	627	7	CN119882	CN119882 EC0CAA002
	74	16.8	80.0	637	4	BG099869	BG099869 uy13c02.y
	75	16.8	80.0	641	6	CD766336	CD766336 AGENCOURT
	76	16.8	80.0	642	1	AL863627	AL863627 AL863627
	77	16.8	80.0	642	1	AL878478	AL878478 AL878478
	78	16.8	80.0	649	7	CF732690	CF732690 UI-M-HA0-
	79	16.8	80.0	651	5	BX315194	BX315194 BX315194
	80	16.8	80.0	655	6	CD772714	CD772714 AGENCOURT
c	81	16.8	80.0	660	8	AZ428022	AZ428022 1M0210N17
c	82	16.8	80.0	679	4	BG381620	BG381620 UI-R-CT0-
	83	16.8	80.0	685	9	AG151855	AG151855 Pan trogl
	84	16.8	80.0	688	5	BQ746626	BQ746626 UI-M-ER0-
	85	16.8	80.0	701	7	CK654484	CK654484 AGENCOURT
c	86	16.8	80.0	720	8	AZ247538	AZ247538 RPCI-23-9
c	87	16.8	80.0	720	8	AZ247542	AZ247542 RPCI-23-9
	88	16.8	80.0	734	7	CF148724	CF148724 AGENCOURT
	89	16.8	80.0	734	7	CO806312	CO806312 AGENCOURT
	90	16.8	80.0	738	6	CB598900	CB598900 AGENCOURT
c	91	16.8	80.0	744	9	AG527299	AG527299 Mus muscu
	92	16.8	80.0	749	7	CO562201	CO562201 AGENCOURT
	93	16.8	80.0	753	7	CR441613	CR441613 CR441613
	94	16.8	80.0	765	7	CF149901	CF149901 AGENCOURT
	95	16.8	80.0	768	9	CC553492	CC553492 CH240_459
	96	16.8	80.0	822	9	AY413996	AY413996 Mus muscu
	97	16.8	80.0	829	5	BU412600	BU412600 602954786

```
    98    16.8   80.0    835  4  BI684946                  BI684946 603307795
c   99    16.8   80.0    844  9  CR017849                  CR017849 Forward s
   100    16.8   80.0    847  7  CK793898                  CK793898 AGENCOURT
```

ALIGNMENTS

```
RESULT 1
CNS04CAJ
LOCUS       CNS04CAJ                1041 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            099K23 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL284212
VERSION     AL284212.1  GI:8022590
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 1041)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="099K23"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG099AF12LP1~end : T7"
ORIGIN

  Query Match             90.5%;  Score 19;  DB 9;  Length 1041;
  Best Local Similarity   90.5%;  Pred. No. 2e+02;
  Matches   19;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              :|||||||||||||||||| ||
Db        942 SAGTGACATGCAGGTCTACCT 962


RESULT 2
CD500170
LOCUS       CD500170                1157 bp    mRNA    linear   EST 12-JUN-2003
DEFINITION  CDA43-E06.x1d-t SHGC-CDA Gasterosteus aculeatus cDNA clone
            CDA43-E06 5', mRNA sequence.
ACCESSION   CD500170
VERSION     CD500170.1  GI:31427201
KEYWORDS    EST.
SOURCE      Gasterosteus aculeatus (three spined stickleback)
  ORGANISM  Gasterosteus aculeatus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
            Gasterosteidae; Gasterosteus.
REFERENCE   1  (bases 1 to 1157)
  AUTHORS   Kingsley,D.M., Peichel,C., Balabahdra,S., Grimwood,J., Dickson,M.,
            Schmutz,J. and Myers,R.M.
  TITLE     Expressed sequence tags from Gasterosteus aculeatus
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Kingsley, DM
            HHMI and Department of Developmental Biology
            Stanford University School of Medicine
            Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA
            Tel: 650 725 5954
            Fax: 650 725 7739
            Email: kingsley@cmgm.stanford.edu
            Plate: 43
            High quality sequence stop: 785.
FEATURES             Location/Qualifiers
     source          1. .1157
                     /organism="Gasterosteus aculeatus"
                     /mol_type="mRNA"
                     /strain="Salinas river, CA"
                     /db_xref="taxon:69293"
                     /clone="CDA43-E06"
                     /sex="mixed male and female"
                     /tissue_type="heads and internal organs combined"
                     /dev_stage="adult"
                     /clone_lib="SHGC-CDA"
                     /note="Vector: lambda ZAP Express/pBK-CMV; Site_1: EcoR1
                     (5' adaptor); Site_2: XhoI (3' linker primer); The mixed
                     organ cDNA library was generated using the ZAP-cDNA method
                     by Stratagene. First strand cDNA synthesis was primed with
                     a a 50 bp linker primer containing an oligo dT sequence
                     preceeded by a synthetic XhoI site. 5 prime adaptors were
                     used containing an EcoR1 cohesive end. The finished cDNAs
                     were inserted in to the ZAP express vector
                     unidirectionally in the sense orientation with respect to
                     the lacZ promoter of pBK-CMV. An amplified library was
                     prepared from approximately 3 million primary clones in
                     the lambda ZAP Express vector. In vivo excision was then
                     used to generate individual pBK-CMV phagemid clones for
                     EST sequencing."
ORIGIN

  Query Match             87.6%;  Score 18.4;  DB 6;  Length 1157;
  Best Local Similarity   95.0%;  Pred. No. 4e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 AGTGACATGCAGGTCTAGCT 21
              |||||||||||||||||| ||
Db        132 AGTGACATGCAGGTCTACCT 151


RESULT 3
AZ995872
LOCUS       AZ995872                 211 bp    DNA     linear   GSS 27-APR-2001
DEFINITION  2M0281B23R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
            clone UUGC2M0281B23 R, genomic survey sequence.
ACCESSION   AZ995872
VERSION     AZ995872.1  GI:13867099
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
REFERENCE   1  (bases 1 to 211)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
            Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
            Niederhausern,A. and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0281  row: B  column: 23
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 211.
FEATURES             Location/Qualifiers
     source          1. .211
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0281B23"
                     /sex="Female"
                     /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC2M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (female) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 8;  Length 211;
  Best Local Similarity   90.5%;  Pred. No. 6.1e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              ||| | ||||||||||||||||
Db        136 CAGAGCCATGCAGGTCTAGCT 156


RESULT 4
AZ218630/c
LOCUS       AZ218630                611 bp    DNA     linear   GSS 09-JUN-2000
DEFINITION  Sheared DNA-82B8.TF Sheared DNA Trypanosoma brucei genomic clone
            Sheared DNA-82B8, genomic survey sequence.
ACCESSION   AZ218630
VERSION     AZ218630.1  GI:8436430
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
```

```
REFERENCE   1  (bases 1 to 611)
  AUTHORS   El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
            Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S.,
            Donelson,J., Fraser,C. and Adams,M.
  TITLE     Determination of clone end sequences from Trypanosoma brucei GUTat
            10.1 sheared DNA library
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: Sheared DNA-82B8.TR
            Contact: Najib M. El-Sayed
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: nelsayed@tigr.org
            Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
            DNA library constructed at TIGR. Clones will be available for
            distribution through Research Genetics, Alabama, USA. Sheared DNA
            end sequences search page: http://www.tigr.org/tdb/mdb/tbdb/.
            Seq primer: M13-Forward
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .611
                     /organism="Trypanosoma brucei"
                     /mol_type="genomic DNA"
                     /strain="TREU927/4 GUTat 10.1"
                     /db_xref="taxon:5691"
                     /clone="Sheared DNA-82B8"
                     /clone_lib="Sheared DNA"
                     /note="Vector: pUC18; Site_1: SmaI; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from a cloned population of
                     Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
                     sheared to give a tight size distribution (approx 2 kb).
                     The v + i method used for the library construction is
                     described in detail in Smith, H.O. and Venter, J.C.
                     (Making small insert libraries for whole genome shotgun
                     sequencing projects.  In Genome Sequencing: A Practical
                     Approach, eds. M. Vaudin and B. Barell, Oxford University
                     Press, 1999)."
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 8;  Length 611;
  Best Local Similarity   90.5%;  Pred. No. 7.2e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              ||||| |||||||||||||| |
Db        548 CAGTGGCATGCAGGTCTAGTT 528


RESULT 5
CR794665
LOCUS       CR794665                878 bp    DNA     linear   GSS 24-SEP-2004
DEFINITION  GR0AAA12AC07RM1 INRA BAC Bos taurus genomic clone INRAb_225E12, DNA
            sequence, genomic survey sequence.
ACCESSION   CR794665
VERSION     CR794665.1  GI:52675664
KEYWORDS    GSS.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE   1  (bases 1 to 878)
  AUTHORS   Eggen,A., Schibler,L. and Roy,A.
  TITLE     Bovine BAC End Sequences from the INRA bovine BAC library
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 878)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-SEP-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
```

```
            - Web : www.genoscope.cns.fr)
COMMENT     Contact: Andre Eggen
            Department of Animal Genetics - LGbC
            INRA
            78350 Jouy-en-Josas, France
            Tel: 33 1 34 65 24 24
            Fax: 33 1 34 65 24 78
            Email: eggen@jouy.inra.fr
            Clones are derived from the INRA bovine BAC library
            (http://locus.jouy.inra.fr/fpc/cattle_bac_map.htm). For BAC library
            availability, please contact Andre Eggen (eggen@jouy.inra.fr). This
            work  was undertaken as part of the International Bovine BAC
            Mapping Consortium (IBBMC) by INRA (Jouy-en-Josas) and Genoscope
            (Evry) Plate: 225 row: E column: 12
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .878
                      /organism="Bos taurus"
                      /mol_type="genomic DNA"
                      /strain="breed: Holstein"
                      /db_xref="taxon:9913"
                      /clone="INRAb_225E12"
                      /sex="Male"
                      /cell_type="fibroblast"
                      /clone_lib="INRA bovine BAC"
                      /note="Vector: pBeloBAC11; Site_1: HindIII; Holstein bull;
                      INRA Bovine BAC library (Male) produced by Andre
                      Eggen~Genoscope sequence ID : GR0AAA12AC07RM1"
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 9;  Length 878;
  Best Local Similarity   90.5%;  Pred. No. 7.7e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              |||| ||| ||||||||||||
Db        297 CAGTCACAAGCAGGTCTAGCT 317


RESULT 6
AK083880/c
LOCUS       AK083880               4261 bp    mRNA    linear   HTC 03-APR-2004
DEFINITION  Mus musculus 12 days embryo spinal ganglion cDNA, RIKEN full-length
            enriched library, clone:D130043C18 product:unclassifiable, full
            insert sequence.
ACCESSION   AK083880
VERSION     AK083880.1  GI:26101557
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 4261)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
            URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.jp/
            URL:http://fantom.gsc.riken.jp/.
FEATURES              Location/Qualifiers
     source           1. .4261
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /strain="C57BL/6J"
                      /db_xref="FANTOM_DB:D130043C18"
                      /db_xref="taxon:10090"
                      /clone="D130043C18"
                      /tissue_type="spinal ganglion"
                      /clone_lib="RIKEN full-length enriched mouse cDNA library"
                      /dev_stage="12 days embryo"
     misc_feature     1. .4261
                      /note="unclassifiable"
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 3;  Length 4261;
  Best Local Similarity   90.5%;  Pred. No. 1e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              |||||| |||||||| |||||
Db       1497 CAGTGATATGCAGGTGTAGCT 1477


RESULT 7
CL289600/c
```

```
LOCUS       CL289600                380 bp    DNA     linear   GSS 10-FEB-2004
DEFINITION  ZMMBBb0641H10r ZMMBBb (HindIII) Zea mays genomic clone
            ZMMBBb0641H10 3', genomic survey sequence.
ACCESSION   CL289600
VERSION     CL289600.1  GI:42503987
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 380)
  AUTHORS   Bharti,A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C.,
            Zohovetz,V., Fuks,G., Yu,Y., Wing,R. and Messing,J.
  TITLE     Sequencing of the maize genome at PGIR (2003c)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Bharti,A.K.
            Dr.Joachim Messing's lab
            The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers
            University
            190 Frelinghuysen Road, Piscataway, NJ 08854, USA
            Tel: 732 445 3801
            Fax: 732 445 5735
            Email: bharti@waksman.rutgers.edu
            Seq primer: SP6
            Class: BAC ends
            High quality sequence start: 99.
FEATURES             Location/Qualifiers
     source          1. .380
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBBb0641H10"
                     /lab_host="E. coli DH10B"
                     /clone_lib="ZMMBBb (HindIII)"
                     /note="Vector: pCUGI; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             82.9%;  Score 17.4;  DB 9;  Length 380;
  Best Local Similarity   94.7%;  Pred. No. 1.1e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAG 19
              || |||||||||||||||||
Db        210 CATTGACATGCAGGTCTAG 192


RESULT 8
CG176005
LOCUS       CG176005                718 bp    DNA     linear   GSS 21-AUG-2003
DEFINITION  PUIMY95TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa0620O21,
            genomic survey sequence.
ACCESSION   CG176005
VERSION     CG176005.1  GI:34066803
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 718)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
            Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: PUIMY95TB
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .718
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa0620O21"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
ORIGIN

  Query Match             82.9%;  Score 17.4;  DB 9;  Length 718;
  Best Local Similarity   94.7%;  Pred. No. 1.2e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAG 19
              || |||||||||||||||||
Db        374 CATTGACATGCAGGTCTAG 392


RESULT 9
BF607360
LOCUS       BF607360                722 bp    mRNA    linear   EST 01-APR-2001
DEFINITION  MY1_000302 Mouse 9-day fetus cDNA library ICRFp522 Mus musculus
            cDNA clone ICRFp522J2149 5', mRNA sequence.
ACCESSION   BF607360
VERSION     BF607360.1  GI:13503852
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 722)
  AUTHORS   Yahyawi,M., Hennig,S., Neidhardt,L., Radelof,U., Hermann,B.G.,
            Lehrach,H. and O'Brien,J.
  TITLE     Detection of a high number of novel genes in a 9-day mouse embryo
            cDNA library normalised by oligonucleotide fingerprinting
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Hennig S
            laboraty 123, dept.Lehrach
            Max-Planck-Institut fuer Molekulare Genetik
            Ihnestr.63-73, D-14195 Berlin, Germany
            Tel: +49 30 8413 1612
            Fax: +49 30 8413 1380
            Email: hennig@molgen.mpg.de
            EST's are made from clones being representatives of clone clusters.
            Clone clusters were calculated from oligonucleotide fingerprints.
            PCR PRimers
            FORWARD: 5'-GAGCTATTCCAGAAGTAGTGA-3'
            BACKWARD: 5'-TAATACGACTCACTATAGGG-3'
            Seq primer: 5'-ATTTAGGTGACACTATAG-3'
            High quality sequence stop: 722.
FEATURES             Location/Qualifiers
     source          1. .722
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="ICRFp522J2149"
                     /tissue_type="whole embryo"
                     /dev_stage="embryonic 9-day"
                     /lab_host="E.coli, XL1 blue"
                     /clone_lib="Mouse 9-day fetus cDNA library ICRFp522"
                     /note="Vector: PSVSport1; Site_1: NotI; Site_2: SalI;
                     Library preparation by oligo_dT priming of RNA. Clones can
                     be ordered from the Resource Center in Berlin,
                     http://www.rzpd.de."
ORIGIN

  Query Match             82.9%;  Score 17.4;  DB 2;  Length 722;
```

```
  Best Local Similarity   90.0%;  Pred. No. 1.2e+03;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGC 20
              ||||  ||||||||||||||
Db        314 CAGTANCATGCAGGTCTAGC 333


RESULT 10
CC340825/c
LOCUS       CC340825                772 bp    DNA     linear   GSS 16-MAY-2003
DEFINITION  OGQAQ84TH ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0368M23,
            genomic survey sequence.
ACCESSION   CC340825
VERSION     CC340825.1  GI:30810231
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 772)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  TITLE     Consortium for Maize Genomics
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .772
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBMa0368M23"
                     /clone_lib="ZM_0.7_1.5_KB"
                     /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
                     methylation filtered genomic DNA library"
ORIGIN

  Query Match             82.9%;  Score 17.4;  DB 8;  Length 772;
  Best Local Similarity   94.7%;  Pred. No. 1.2e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAG 19
              || |||||||||||||||
Db        744 CATTGACATGCAGGTCTAG 726


RESULT 11
CG211386
LOCUS       CG211386                772 bp    DNA     linear   GSS 22-AUG-2003
DEFINITION  OG5CS03TC ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0832B05,
            genomic survey sequence.
ACCESSION   CG211386
VERSION     CG211386.1  GI:34111216
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 772)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  TITLE     Consortium for Maize Genomics
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .772
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBMa0832B05"
                     /clone_lib="ZM_0.7_1.5_KB"
                     /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
                     methylation filtered genomic DNA library"
ORIGIN

  Query Match             82.9%;  Score 17.4;  DB 9;  Length 772;
  Best Local Similarity   94.7%;  Pred. No. 1.2e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAG 19
              || |||||||||||||||
Db        287 CATTGACATGCAGGTCTAG 305


RESULT 12
CG362883/c
LOCUS       CG362883                805 bp    DNA     linear   GSS 26-AUG-2003
DEFINITION  OGYBS24TH ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0643C23,
            genomic survey sequence.
ACCESSION   CG362883
VERSION     CG362883.1  GI:34280150
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 805)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  TITLE     Consortium for Maize Genomics
  JOURNAL   Unpublished (2002)
COMMENT     Other_GSSs: OGYBS24TV
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .805
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBMa0643C23"
                     /clone_lib="ZM_0.7_1.5_KB"
                     /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
                     methylation filtered genomic DNA library"
ORIGIN

  Query Match             82.9%;  Score 17.4;  DB 9;  Length 805;
  Best Local Similarity   94.7%;  Pred. No. 1.2e+03;
```

```
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAG 19
              || |||||||||||||||||
Db        384 CATTGACATGCAGGTCTAG 366


RESULT 13
BI548151/c
LOCUS       BI548151                816 bp    mRNA    linear   EST 05-SEP-2001
DEFINITION  603189492F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5260847 5',
            mRNA sequence.
ACCESSION   BI548151
VERSION     BI548151.1  GI:15435463
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 816)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11657  row: g  column: 24
            High quality sequence stop: 740.
FEATURES             Location/Qualifiers
     source          1. .816
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5260847"
                     /tissue_type="hippocampus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_95"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
                     (gtcgag); Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTTTTVN-3', size-selected for average
                     insert size 2.5 kb and normalized to ROT 5. This is a
                     primary library enriched for full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation). Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health). Note: this
                     is a NIH_MGC Library."
ORIGIN

  Query Match             82.9%;  Score 17.4;  DB 4;  Length 816;
  Best Local Similarity   94.7%;  Pred. No. 1.2e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAG 19
              |||||||| ||||||||||
Db        295 CAGTGACAGGCAGGTCTAG 277


RESULT 14
CC340835
LOCUS       CC340835                817 bp    DNA     linear   GSS 16-MAY-2003
DEFINITION  OGQAQ84TV ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0368M23,
            genomic survey sequence.
ACCESSION   CC340835
VERSION     CC340835.1  GI:30810241
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 817)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  TITLE     Consortium for Maize Genomics
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .817
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBMa0368M23"
                     /clone_lib="ZM_0.7_1.5_KB"
                     /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
                     methylation filtered genomic DNA library"
ORIGIN

  Query Match             82.9%;  Score 17.4;  DB 8;  Length 817;
  Best Local Similarity   94.7%;  Pred. No. 1.2e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAG 19
              || |||||||||||||||||
Db        488 CATTGACATGCAGGTCTAG 506


RESULT 15
CG211861/c
LOCUS       CG211861                923 bp    DNA     linear   GSS 22-AUG-2003
DEFINITION  OG1AW29TV ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0720F09,
            genomic survey sequence.
ACCESSION   CG211861
VERSION     CG211861.1  GI:34111691
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 923)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  TITLE     Consortium for Maize Genomics
  JOURNAL   Unpublished (2002)
COMMENT     Other_GSSs: OG1AW29TH
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .923
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
```

/db_xref="taxon:4577"
/clone="ZMMBMa0720F09"
/clone_lib="ZM_0.7_1.5_KB"
/note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb methylation filtered genomic DNA library"
ORIGIN

Query Match 82.9%; Score 17.4; DB 9; Length 923;
Best Local Similarity 94.7%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAGTGACATGCAGGTCTAG 19
|| ||||||||||||||||
Db 157 CATTGACATGCAGGTCTAG 139

RESULT 16
BY609340
LOCUS BY609340 382 bp mRNA linear EST 15-DEC-2002
DEFINITION BY609340 RIKEN full-length enriched, visual cortex Mus musculus cDNA clone K230308M20 3', mRNA sequence.
ACCESSION BY609340
VERSION BY609340.1 GI:26944522
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 382)
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
MEDLINE 22354683
PUBMED 12466851
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Tissues were provided by Michela Fagiolini and Takao K. Hensch (Laboratory for Neuronal Circuit Development Brain Science Institute RIKEN 2-1 Hirosawa,Wako-shi,Saitama 351-0198 Japan) whose assistance we gratefully acknowledge.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .382
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="K230308M20"
/tissue_type="visual cortex"
/clone_lib="RIKEN full-length enriched, visual cortex"
ORIGIN

Query Match 81.0%; Score 17; DB 6; Length 382;
Best Local Similarity 100.0%; Pred. No. 1.7e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 AGTGACATGCAGGTCTA 18
|||||||||||||||||
Db 28 AGTGACATGCAGGTCTA 44

RESULT 17
BY649695
LOCUS BY649695 408 bp mRNA linear EST 16-DEC-2002
DEFINITION BY649695 RIKEN full-length enriched, visual cortex Mus musculus cDNA clone K530043P16 3', mRNA sequence.
ACCESSION BY649695
VERSION BY649695.1 GI:27005959
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 408)
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,

```
            Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
            Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
            Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
            Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
            Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
            Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
            Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
            Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
            Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
            Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
            Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
  MEDLINE   22354683
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
            Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
            Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,
            Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
            Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
            Hayashizaki,Y. Direct Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
             Tissues were provided by Michela Fagiolini and Takao K. Hensch (
            Laboratory for Neuronal Circuit Development Brain Science Institute
            RIKEN 2-1 Hirosawa,Wako-shi,Saitama 351-0198 Japan ) whose
            assistance we gratefully acknowledge.
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .408
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="K530043P16"
                     /tissue_type="visual cortex"
                     /clone_lib="RIKEN full-length enriched, visual cortex"
ORIGIN

  Query Match             81.0%;  Score 17;  DB 6;  Length 408;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+03;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 AGTGACATGCAGGTCTA 18
              |||||||||||||||||
Db         56 AGTGACATGCAGGTCTA 72


RESULT 18
CA124343/c
LOCUS       CA124343                622 bp    mRNA    linear   EST 23-SEP-2003
DEFINITION  SCQGLR1086F09.g LR1 Saccharum officinarum cDNA clone SCQGLR1086F09
            5', mRNA sequence.
ACCESSION   CA124343
VERSION     CA124343.1  GI:34977651
KEYWORDS    EST.
SOURCE      Saccharum officinarum
  ORGANISM  Saccharum officinarum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Saccharum; Saccharum officinarum
            complex.
REFERENCE   1  (bases 1 to 622)
  AUTHORS   Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
  TITLE     The libraries that made SUCEST
  JOURNAL   Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT     Contact: Arruda P
            Centro de Biologia Molecular e Engenharia Genetica
            Universidade Estadual de Campinas
            Caixa Postal 6010, 13083-970, Campinas SP, Brazil
            Tel: 55 19 3788 1137
            Fax: 55 19 3788 1089
            Email: parruda@unicamp.br
            Clone distribution: clone distribution information can be found
            through the Brazilian Clone Collection Center (BCCC) at
            http://www.bcccenter.fcav.unesp.br
            Plate: 086  row: F  column: 09
            Seq primer: T7 Promoter Primer.
FEATURES             Location/Qualifiers
     source          1. .622
                     /organism="Saccharum officinarum"
                     /mol_type="mRNA"
                     /db_xref="taxon:4547"
                     /clone="SCQGLR1086F09"
                     /lab_host="DH10B"
                     /clone_lib="LR1"
                     /note="Organ: Leaf roll from field grown adult plants
                     (large insert library); Vector: pSport1; Site_1: SalI;
                     Site_2: NotI; An unidirectional cDNA library generated
                     from [Leaf roll from field grown  adult plants (large
                     insert library)]. cDNA was prepared from polyA+ mRNA
                     using SuperScript Plasmid System Kit (Invitrogen). The
                     double-strand  cDNAs were fractionated in a sepharose
                     CL-2B 40cm-columns and fragments  sizing between 0.8 and
                     1.5 Kb were directionally cloned into the vector.  Details
                     of each source of RNA and library construction can be
                     obtained at  http://sucest.lad.ic.unicamp.br/public"
ORIGIN

  Query Match             81.0%;  Score 17;  DB 6;  Length 622;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+03;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 TGACATGCAGGTCTAGC 20
              |||||||||||||||||
Db        444 TGACATGCAGGTCTAGC 428


RESULT 19
AZ069535
LOCUS       AZ069535                633 bp    DNA     linear   GSS 30-MAR-2000
DEFINITION  RPCI-23-435G22.TJ RPCI-23 Mus musculus genomic clone
            RPCI-23-435G22, genomic survey sequence.
ACCESSION   AZ069535
VERSION     AZ069535.1  GI:7360787
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 633)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S.,
```

```
            Akinret,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de
            Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-23-435G22.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 435  row: G  column: 22
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .633
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-435G22"
                     /sex="Female"
                     /lab_host="DH10B"
                     /clone_lib="RPCI-23"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).   "
ORIGIN

  Query Match             81.0%;  Score 17;  DB 8;  Length 633;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+03;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCT 17
              |||||||||||||||||
Db        122 CAGTGACATGCAGGTCT 138


RESULT 20
BB650662/c
LOCUS       BB650662                642 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB650662 RIKEN full-length enriched, 0 day neonate cerebellum Mus
            musculus cDNA clone C230020D16 5', mRNA sequence.
ACCESSION   BB650662
VERSION     BB650662.1  GI:16484917
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 642)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
            Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,
            Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,
            Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
            and Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
            Sugahara,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I.,
            Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .642
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="C230020D16"
                     /tissue_type="cerebellum"
                     /dev_stage="0 day neonate"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, 0 day neonate
                     cerebellum"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 20.0 and subtraction to Rot = 479.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I."
ORIGIN

  Query Match             81.0%;  Score 17;  DB 2;  Length 642;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+03;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCT 17
              |||||||||||||||||
Db         27 CAGTGACATGCAGGTCT 11


RESULT 21
BB293162/c
LOCUS       BB293162                655 bp    mRNA    linear   EST 24-OCT-2001
```

```
DEFINITION  BB293162 RIKEN full-length enriched, 9.5 days embryo parthenogenote
            Mus musculus cDNA clone B130017A20 3' similar to AK001617 Homo
            sapiens cDNA FLJ10755 fis, clone NT2RP3004569, weakly similar to
            ANKYRIN, BRAIN VARIANT 1, mRNA sequence.
ACCESSION   BB293162
VERSION     BB293162.2  GI:16401650
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 655)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
            Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,
            Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,
            Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On Jul 10, 2000 this sequence version replaced gi:8993654.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
            and Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
            Sugahara,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I.,
            Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .655
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="B130017A20"
                     /tissue_type="parthenogenote"
                     /dev_stage="9.5 days embryo"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, 9.5 days embryo
                     parthenogenote"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of subtraction to
                     Rot = 229.0 Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA
                     was cleaved with XhoI and BamHI. Vector: a modified
                     pBluescript KS(+) after bulk excision from Lambda FLC I."
ORIGIN

  Query Match             81.0%;  Score 17;  DB 2;  Length 655;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+03;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCT 17
              |||||||||||||||||
Db        145 CAGTGACATGCAGGTCT 129


RESULT 22
AG457137
LOCUS       AG457137                768 bp    DNA     linear   GSS 04-JUN-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-343P07.TJ, genomic survey
            sequence.
ACCESSION   AG457137
VERSION     AG457137.1  GI:48148651
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     BAC end Sequences of Library MSMg01
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 768)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .768
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-343P07.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN
```

```
  Query Match             81.0%;  Score 17;  DB 9;  Length 768;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+03;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCT 17
              |||||||||||||||||
Db        191 CAGTGACATGCAGGTCT 207


RESULT 23
AG557182
LOCUS       AG557182                781 bp    DNA     linear   GSS 05-JUN-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-475F07.T7, genomic survey
            sequence.
ACCESSION   AG557182
VERSION     AG557182.1  GI:48317880
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     BAC end Sequences of Library MSMg01
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 781)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : T7
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .781
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-475F07.T7"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match             81.0%;  Score 17;  DB 9;  Length 781;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+03;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCT 17
              |||||||||||||||||
Db        187 CAGTGACATGCAGGTCT 203


RESULT 24
CF411086
LOCUS       CF411086                949 bp    mRNA    linear   EST 02-SEP-2003
DEFINITION  CH3#071_D10MF Canine heart normalized cDNA Library in pBluescript
            Canis familiaris cDNA clone CH3#071_D10 5', mRNA sequence.
ACCESSION   CF411086
VERSION     CF411086.1  GI:34412332
KEYWORDS    EST.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1  (bases 1 to 949)
  AUTHORS   Yi,Y., Desai,R., Olarte,M., Henthorn,P. and George A.L.
  TITLE     Expressed sequence tags from Canine heart
  JOURNAL   Unpublished (2003)
COMMENT     Other_ESTs: CH3#071_D10MR
            Contact: George AL
            Division of Genetic Medicine
            Vanderbilt University
            529 Light Hall, 2215 Garland Avenue, Nashville, TN 37232-0275, USA
            Tel: 615 936 2660
            Fax: 615 936 2661
            Email: al.george@vanderbilt.edu
            Insert Length: 1180   Std Error: 0.00
            Seq primer: MF: GTTTTCCCAGTCACGACGTTG
            High quality sequence start: 221
            High quality sequence stop: 683.
FEATURES             Location/Qualifiers
     source          1. .949
                     /organism="Canis familiaris"
                     /mol_type="mRNA"
                     /db_xref="taxon:9615"
                     /clone="CH3#071_D10"
                     /tissue_type="heart"
                     /cell_type="heart"
                     /dev_stage="mixed developmental stages (adult, 30 day - 40
                     day fetal)"
                     /clone_lib="Canine heart normalized cDNA Library in
                     pBluescript"
                     /note="Organ: heart; Vector: pBluescript; Site_1: 5' of
                     vector NotI; Site_2: 3' of vector EcoRI; Tissue source:
                     dog heart (adult, 30 day - 40 day fetal), right and left
                     atria and ventricle. Dog breed - mixed (beagle, German
                     shepherd,  pointer, Irish setter). Library construction:
                     oligo-dT primed"
ORIGIN

  Query Match             81.0%;  Score 17;  DB 7;  Length 949;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+03;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCT 17
              |||||||||||||||||
Db        299 CAGTGACATGCAGGTCT 315


RESULT 25
AK044974/c
LOCUS       AK044974                2445 bp    mRNA    linear   HTC 03-APR-2004
DEFINITION  Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length
            enriched library, clone:B130017A20 product:synuclein, alpha
            interacting protein (synphilin), full insert sequence.
ACCESSION   AK044974
VERSION     AK044974.1  GI:26336964
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
```

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 2445)
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Tissues were provided by Dr. Tomohiro Kono (Department of Animal Science, Tokyo University of Agriculture, 1737 Hunako Atsugi City, Kanagawa Prefecture, Japan) whose assistance we gratefully acknowledge.
Please visit our web site for further details.
URL:http://genome.gsc.riken.jp/
URL:http://fantom.gsc.riken.jp/.
FEATURES Location/Qualifiers
source 1. .2445
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:B130017A20"
/db_xref="taxon:10090"
/clone="B130017A20"
/tissue_type="parthenogenote"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="9.5 days embryo"
CDS 385. .>2445
/note="unnamed protein product; putative synuclein, alpha interacting protein (synphilin) (MGD|MGI:1915097, GB|NM_026408, evidence: BLASTN, 99%, match=2564)"
/codon_start=1
/protein_id="BAC32166.1"
/db_xref="GI:26336965"
/translation="MEAPEYLDLDEIDFSDDISYSVTSLKTIPALCRRCDSQNEDRSV
SSSGWNCGVSTLITNPQKPTGIADVYSKFRPVKRVSPLKHQPETLENNENEDQKNNTV
EYQKGGETDQGPQPEELSPEDGVGGLPGKGSEPSQALGELEHYDLDMDEILDVPYIKS
SQQLAPLTKVTSEKRILGLCTTINGLSAKTCPIASTENSTPNMTPFCVLSPVKSPHLR
KAPTALRDQHKLSTEDSESSPALGKCGPAYESENHSKDFLNKVFSDPHSRKIEKSGPD
CKLRSFHLQSSAAGAKTEEPINGMNWTNTQGTEERTEYLKKVRSILNIVNEGQISLLP
HLAADNLDKIHDENGNNLLHIAASKGHAECLQHLTSLMGEDCLNERNTEQLTPAGLAI
KNGQLECVRWMVSETEAIAELSCSKDFPSLIHYAGCYGQEKILLWLLQFMQEQGISLD
EVDREGNSAVHVASQHGYLGCIQTLVEYGANVTMQNHAGEKPSQSAERHGHTLCSRYL
VVVETCMSLASQVVKLTKQLKEQTVERVTLQSQLQQLLEAQKSEGKSLPSSPSSPSSP
ASTKSQWKALDTDEESTGKSKVGAQEGIQVLGSLSVSSRARTKGKDEDSDKILRQLLG
KEISENVCTQEKLSLEFQDAQASCRNSKKIPLEKRELKLARLRQLMQRSLSESDTDSN
NSEDP"
ORIGIN

Query Match 81.0%; Score 17; DB 3; Length 2445;
Best Local Similarity 100.0%; Pred. No. 2.3e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAGTGACATGCAGGTCT 17
||||||||||||||||
Db 1935 CAGTGACATGCAGGTCT 1919

Search completed: September 6, 2005, 21:56:00
Job time : 1506.84 secs

OM nucleic - nucleic search, using sw model

Run on: September 6, 2005, 16:01:23 ; Search time 189.656 Seconds
(without alignments)
655.473 Million cell updates/sec

Title: US-10-729-421-40
Perfect score: 21
Sequence: 1 cagtgacatgcaggtctagct 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	21	100.0	21	8	ACC80536	Acc80536 Exemplary
	2	21	100.0	21	8	ACC80538	Acc80538 Internal
c	3	21	100.0	21	8	ACC80537	Acc80537 Internal
c	4	21	100.0	21	9	ABZ59632	Abz59632 Parvoviru
	5	21	100.0	21	12	ADI53796	Adi53796 HAV inter
	6	21	100.0	23	12	ADI53797	Adi53797 HAV inter
	7	21	100.0	23	12	ADQ30670	Adq30670 West Nile
	8	21	100.0	25	12	ADQ30671	Adq30671 West Nile
	9	21	100.0	681	9	ABZ59634	Abz59634 Exemplary
	10	21	100.0	727	12	ADI53795	Adi53795 HAV inter
	11	21	100.0	967	12	ADQ30647	Adq30647 West Nile
	12	21	100.0	1696	8	ACC80539	Acc80539 Internal
c	13	17.4	82.9	5446	10	ADE71204	Ade71204 Novel hum
c	14	17.4	82.9	28482	8	ABZ73855	Abz73855 Secreted
c	15	17.4	82.9	28482	8	ADA44262	Ada44262 Human sec
c	16	17.4	82.9	32681	8	ABZ74517	Abz74517 Secreted
c	17	17.4	82.9	32681	8	ABZ73854	Abz73854 Secreted
c	18	17.4	82.9	32681	8	ADA98915	Ada98915 Human sec
c	19	17.4	82.9	32681	8	ADA44261	Ada44261 Human sec
c	20	17.4	82.9	32681	8	ADA44519	Ada44519 Human sec
c	21	17.4	82.9	32681	10	ADC20949	Adc20949 Human sec
c	22	17.4	82.9	32681	10	ABZ68053	Abz68053 Human sec
	23	16.8	80.0	320	8	ABX46275	Abx46275 Bovine ES
	24	16.8	80.0	36221	4	AAS00624	Aas00624 Human dea
	25	16.4	78.1	72705	11	ACN45158	Acn45158 Human gen
	26	16.4	78.1	110000	8	ABX16390_5	Continuation (6 of
	27	16.4	78.1	117382	11	ACN44804	Acn44804 Mouse gen
c	28	16.4	78.1	340449	8	AAL52198	Aal52198 Human sec
c	29	16.2	77.1	201	13	ADS40801	Ads40801 Human aut
	30	16.2	77.1	201	13	ADS39530	Ads39530 Human aut
c	31	16.2	77.1	394	4	AAI82251	Aai82251 Human pol
	32	16.2	77.1	473	9	ACH36560	Ach36560 Human end
	33	16.2	77.1	497	10	ADB56505	Adb56505 Toxicity-
	34	16.2	77.1	1487	10	ADC71327	Adc71327 Human col
c	35	16.2	77.1	2048	2	AAQ85985	Aaq85985 Zea mays
	36	16.2	77.1	2174	10	ADE62245	Ade62245 Rat gene
c	37	16.2	77.1	2238	5	AAS76384	Aas76384 DNA encod
c	38	16.2	77.1	2569	13	ADR06803	Adr06803 Full leng
c	39	16.2	77.1	2721	8	AAL53547	Aal53547 cDNA of h
	40	16.2	77.1	4445	6	ABA01096	Aba01096 Brevibact
	41	16.2	77.1	9048	4	AAC90812	Aac90812 B. lactof
	42	16.2	77.1	10500	4	AAL05334	Aal05334 Human rep
	43	16.2	77.1	10500	4	ABL98203	Abl98203 Human tes
c	44	16.2	77.1	14902	13	ADS36489	Ads36489 Human aut
c	45	16.2	77.1	15515	8	AAL53548	Aal53548 Genomic D
	46	16.2	77.1	55827	8	ACA60949	Aca60949 DNA encod
c	47	16.2	77.1	58337	13	ADS36454	Ads36454 Human aut
	48	16.2	77.1	64423	13	ADS36462	Ads36462 Human aut
	49	16.2	77.1	70372	6	AAL53466	Aal53466 Ras-like
c	50	16.2	77.1	90442	9	ADA03077	Ada03077 Mouse mCG
c	51	16.2	77.1	90442	9	ADA66361	Ada66361 Mouse mCG
c	52	16.2	77.1	90442	10	ADB72815	Adb72815 Mouse mCG
c	53	16.2	77.1	90442	10	ADC26997	Adc26997 Mouse car
c	54	16.2	77.1	90442	11	ADL27155	Adl27155 Mouse gen
	55	16.2	77.1	143306	6	ABK49586	Abk49586 Human tra
c	56	16.2	77.1	349980	5	AAH68529	Aah68529 C glutami
c	57	15.8	75.2	171	2	AAV89101	Aav89101 EST clone
c	58	15.8	75.2	279	2	AAV87960	Aav87960 EST clone
c	59	15.8	75.2	297	9	ADB08791	Adb08791 Alloiococ
c	60	15.8	75.2	297	9	ADB08789	Adb08789 Alloiococ
	61	15.8	75.2	497	5	AAS88145	Aas88145 DNA encod
	62	15.8	75.2	497	5	AAS80088	Aas80088 DNA encod
c	63	15.8	75.2	587	4	AAH07183	Aah07183 Human cDN
	64	15.8	75.2	748	4	AAH03978	Aah03978 Human cDN
c	65	15.8	75.2	882	10	ADC08897	Adc08897 Rice DNA
	66	15.8	75.2	1614	9	ADB08797	Adb08797 Alloiococ
	67	15.8	75.2	1670	4	AAH17173	Aah17173 Human cDN
c	68	15.8	75.2	1777	4	AAH16391	Aah16391 Human cDN
c	69	15.8	75.2	1995	6	ABL88382	Abl88382 Pain regu
c	70	15.8	75.2	2378	5	AAS64699	Aas64699 DNA encod
c	71	15.8	75.2	2379	5	AAS67041	Aas67041 DNA encod
	72	15.8	75.2	4015	12	ADJ34728	Adj34728 Rat 2'-5'
	73	15.8	75.2	4708	12	ADJ34707	Adj34707 Mouse 2'-
c	74	15.8	75.2	31236	9	ADA02900	Ada02900 Human PTP
c	75	15.8	75.2	31236	10	ADB72638	Adb72638 Human PTP
c	76	15.8	75.2	31236	10	ADC85379	Adc85379 Mouse Ptp
c	77	15.8	75.2	31236	12	ADM74495	Adm74495 Human car
	78	15.8	75.2	31718	4	AAK90359	Aak90359 Human dig
	79	15.8	75.2	31718	4	AAK90360	Aak90360 Human dig
c	80	15.8	75.2	31718	4	AAK73104	Aak73104 Human imm
c	81	15.8	75.2	31718	4	AAK87573	Aak87573 Human imm
c	82	15.8	75.2	31718	4	AAK73120	Aak73120 Human imm
	83	15.8	75.2	31718	4	AAK87442	Aak87442 Human imm
	84	15.8	75.2	31718	4	AAK87443	Aak87443 Human imm
c	85	15.8	75.2	31718	4	AAK87592	Aak87592 Human imm
	86	15.8	75.2	31718	4	AAL06415	Aal06415 Human rep
	87	15.8	75.2	31718	4	AAL06416	Aal06416 Human rep
	88	15.8	75.2	31718	5	AAS39916	Aas39916 Genomic s
	89	15.8	75.2	31718	5	AAS39915	Aas39915 Genomic s
	90	15.8	75.2	31718	9	ADB32875	Adb32875 Human nov
	91	15.8	75.2	31718	9	ADB32876	Adb32876 Human nov
c	92	15.8	75.2	31718	12	ADN41665	Adn41665 Novel hum
c	93	15.8	75.2	31718	12	ADN41666	Adn41666 Novel hum

c	94	15.8	75.2	63609	12	ADQ97537 Adq97537 Human can
	95	15.8	75.2	110000	9	ADB12064_07 Continuation (8 of
	96	15.8	75.2	122923	11	ACN44026 Acn44026 Human gen
c	97	15.8	75.2	170170	10	ADL13643 Adl13643 Osteoarth
	98	15.4	73.3	20	6	AAD34672 Aad34672 DST CHS1_
c	99	15.4	73.3	327	6	AAD34632 Aad34632 HBV infec
	100	15.4	73.3	640	12	ADJ75740 Adj75740 Marker ge

ALIGNMENTS

RESULT 1
ACC80536/c
ID ACC80536 standard; DNA; 21 BP.
XX
AC ACC80536;
XX
DT 29-AUG-2003 (first entry)
XX
DE Exemplary sequence for method detecting HBV DNA in a sample.
XX
KW Hepatitis B virus; diagnosis; nuclease assay; ss;
KW transcription-mediated amplification.
XX
OS Hepatitis B virus.
XX
PN WO2003031934-A2.
XX
PD 17-APR-2003.
XX
PF 09-OCT-2002; 2002WO-US032367.
XX
PR 09-OCT-2001; 2001US-0328492P.
PR 29-MAR-2002; 2002US-0368823P.
PR 02-JUL-2002; 2002US-0393561P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Shyamala V;
XX
DR WPI; 2003-403124/38.
XX
PT New isolated hepatitis B virus (HBV) capture oligonucleotides, useful for
PT detecting HBV infection in a biological sample, or in capturing HBV
PT nucleic acids.
XX
PS Disclosure; Page 22; 50pp; English.
XX
CC The invention relates to a novel method of detecting hepatitis B virus
CC (HBV) infections in e.g. blood samples from donors, by capturing and
CC amplifying conserved regions of the HBV genome using a transcription-
CC mediated amplification (TMA) method as well as a 5' nuclease assay. This
CC sequence represents an exemplary replacement sequence for a target
CC sequence (ACC80535) used in an internal control for the method of the
CC invention. The new method is very sensitive and is able to detect about
CC 100 infectious units (IU) of HBV in a viremic sample
XX
SQ Sequence 21 BP; 5 A; 6 C; 5 G; 5 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 8; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CAGTGACATGCAGGTCTAGCT 21
              |||||||||||||||||||||
Db         21 CAGTGACATGCAGGTCTAGCT 1
```

RESULT 2
ACC80538
ID ACC80538 standard; DNA; 21 BP.
XX
AC ACC80538;
XX
DT 29-AUG-2003 (first entry)
XX
DE Internal control primer #2 for hepatitis B virus DNA detection method.
XX
KW Primer; PCR; ss; hepatitis B virus; diagnosis; nuclease assay;
KW transcription-mediated amplification.
XX
OS Hepatitis B virus.
XX
FH Key Location/Qualifiers
FT misc_difference 1
FT /*tag= a
FT /note= "linked to 6-FAM"
FT misc_difference 21
FT /*tag= b
FT /note= "linked to TAMRA"
XX
PN WO2003031934-A2.
XX
PD 17-APR-2003.
XX
PF 09-OCT-2002; 2002WO-US032367.
XX
PR 09-OCT-2001; 2001US-0328492P.
PR 29-MAR-2002; 2002US-0368823P.
PR 02-JUL-2002; 2002US-0393561P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Shyamala V;
XX
DR WPI; 2003-403124/38.
XX
PT New isolated hepatitis B virus (HBV) capture oligonucleotides, useful for
PT detecting HBV infection in a biological sample, or in capturing HBV
PT nucleic acids.
XX
PS Claim 14; Fig 2; 50pp; English.
XX
CC The invention relates to a novel method of detecting hepatitis B virus
CC (HBV) infections in e.g. blood samples from donors, by capturing and
CC amplifying conserved regions of the HBV genome using a transcription-
CC mediated amplification (TMA) method as well as a 5' nuclease assay. The
CC method may also use an internal control sequence such as ACC80539, to
CC determine the level of amplification and detection by the primers and
CC probes used in the method of the invention. This sequence represents a
CC primer used to amplify the internal control region DNA sequence. The new
CC method is very sensitive and is able to detect about 100 infectious units
CC (IU) of HBV in a viremic sample
XX
SQ Sequence 21 BP; 5 A; 5 C; 6 G; 5 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 8; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CAGTGACATGCAGGTCTAGCT 21
              |||||||||||||||||||||
Db          1 CAGTGACATGCAGGTCTAGCT 21
```

RESULT 3
ACC80537/c
ID ACC80537 standard; DNA; 21 BP.
XX
AC ACC80537;
XX
DT 29-AUG-2003 (first entry)
XX
DE Internal control primer #1 for hepatitis B virus DNA detection method.
XX

```
KW   Primer; PCR; ss; hepatitis B virus; diagnosis; nuclease assay;
KW   transcription-mediated amplification.
XX
OS   Hepatitis B virus.
XX
FH   Key             Location/Qualifiers
FT   misc_difference 1
FT                   /*tag=  a
FT                   /note= "linked to 6-FAM"
FT   misc_difference 21
FT                   /*tag=  b
FT                   /note= "linked to TAMRA"
XX
PN   WO2003031934-A2.
XX
PD   17-APR-2003.
XX
PF   09-OCT-2002; 2002WO-US032367.
XX
PR   09-OCT-2001; 2001US-0328492P.
PR   29-MAR-2002; 2002US-0368823P.
PR   02-JUL-2002; 2002US-0393561P.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Shyamala V;
XX
DR   WPI; 2003-403124/38.
XX
PT   New isolated hepatitis B virus (HBV) capture oligonucleotides, useful for
PT   detecting HBV infection in a biological sample, or in capturing HBV
PT   nucleic acids.
XX
PS   Claim 13; Fig 2; 50pp; English.
XX
CC   The invention relates to a novel method of detecting hepatitis B virus
CC   (HBV) infections in e.g. blood samples from donors, by capturing and
CC   amplifying conserved regions of the HBV genome using a transcription-
CC   mediated amplification (TMA) method as well as a 5' nuclease assay. The
CC   method may also use an internal control sequence such as ACC80539, to
CC   determine the level of amplification and detection by the primers and
CC   probes used in the method of the invention. This sequence represents a
CC   primer used to amplify the internal control region DNA sequence. The new
CC   method is very sensitive and is able to detect about 100 infectious units
CC   (IU) of HBV in a viremic sample
XX
SQ   Sequence 21 BP; 5 A; 6 C; 5 G; 5 T; 0 U; 0 Other;

  Query Match            100.0%;  Score 21;  DB 8;  Length 21;
  Best Local Similarity  100.0%;  Pred. No. 2.5;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              |||||||||||||||||||||
Db         21 CAGTGACATGCAGGTCTAGCT 1


RESULT 4
ABZ59632/c
ID   ABZ59632 standard; DNA; 21 BP.
XX
AC   ABZ59632;
XX
DT   22-APR-2003  (first entry)
XX
DE   Parvovirus B19 related internal control oligonucleotide SEQ ID NO:90.
XX
KW   Human parvovirus B19; parvovirus B19; infection; virus; blood; plasma;
KW   PCR primer; ss.
XX
OS   B19 virus.
OS   Synthetic.
XX
PN   WO2003002753-A2.
XX
PD   09-JAN-2003.
XX
PF   28-JUN-2002; 2002WO-US020684.
XX
PR   28-JUN-2001; 2001US-0302077P.
PR   19-MAR-2002; 2002US-0365956P.
PR   29-MAR-2002; 2002US-0369224P.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Pichuantes S,  Shyamala V;
XX
DR   WPI; 2003-201510/19.
XX
PT   Detecting a human parvovirus B19 infection in a biological sample to
PT   prevent viral transmission, comprises reacting a parvovirus B19 nucleic
PT   acid with a primer complementary to the 3'-terminal portion of the RNA
PT   target sequence.
XX
PS   Claim 8; Page 29; 148pp; English.
XX
CC   The present invention describes a method for detecting a human parvovirus
CC   B19 infection in a biological sample. The method comprises reacting the
CC   isolated parvovirus B19 nucleic acid with a first oligonucleotide
CC   consisting of a first primer containing a complexing sequence
CC   sufficiently complementary to the 3'-terminal portion of the RNA target
CC   sequence to complex with. Also described: (1) amplifying a target
CC   parvovirus B19 nucleotide sequence; (2) a polynucleotide comprising one
CC   of 47 700 base pair sequences (see ABZ59549 to ABZ59569, and ABZ59604 to
CC   ABZ59629); (3) a polynucleotide comprising either of 2 4678 base pair
CC   sequences (see ABZ59570 and ABZ59571); (4) an oligonucleotide primer
CC   consisting of a promoter region recognised by a DNA-dependent RNA
CC   polymerase operably linked to a human parvovirus B19-specific complexing
CC   sequence of 10-75 nucleotides; (5) an oligonucleotide probe comprising a
CC   parvovirus B19-specific hybridising sequence of 10-50 nucleotides linked
CC   to an acridinium ester label; and (6) a diagnostic test kit comprising an
CC   oligonucleotide primer of (4), and instructions for conducting the
CC   diagnostic test. The method is useful for detecting parvovirus infection
CC   in a biological sample, such as in blood products, to prevent
CC   transmission of the virus through blood and plasma derivatives or by
CC   close personal contact. ABZ59549 to ABZ59634 and ABP57262 to ABP57267
CC   represent sequences used in the exemplification of the present invention
XX
SQ   Sequence 21 BP; 5 A; 6 C; 5 G; 5 T; 0 U; 0 Other;

  Query Match            100.0%;  Score 21;  DB 9;  Length 21;
  Best Local Similarity  100.0%;  Pred. No. 2.5;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              |||||||||||||||||||||
Db         21 CAGTGACATGCAGGTCTAGCT 1


RESULT 5
ADI53796
ID   ADI53796 standard; DNA; 21 BP.
XX
AC   ADI53796;
XX
DT   06-MAY-2004  (first entry)
XX
DE   HAV internal control specific detection probe.
XX
KW   HAV; nucleic acid amplification; nucleic acid detection; ss; probe.
XX
OS   Hepatitis A virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   modified_base   1
```

```
FT                   /*tag=  a
FT                   /mod_base= 5'-TET labelled
FT   modified_base   21
FT                   /*tag=  b
FT                   /mod_base= 3'-TAMRA labelled
XX
PN   WO2003106641-A2.
XX
PD   24-DEC-2003.
XX
PF   12-JUN-2003; 2003WO-US018827.
XX
PR   12-JUN-2002; 2002US-0388544P.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Shyamala V;
XX
DR   WPI; 2004-082181/08.
XX
PT   Hepatitis A virus specific primers and probes derived from conserved
PT   regions of the hepatitis A virus genome, useful in nucleic acid-based
PT   diagnostic tests for the detection of Hepatitis A virus in biological
PT   samples.
XX
PS   Example 2; SEQ ID NO 18; 44pp; English.
XX
CC   The invention relates to Hepatitis A virus (HAV) specific primers and
CC   probes derived from conserved regions of the hepatitis A virus genome.
CC   The HAV-specific primers and probes are used in a method for detecting
CC   HAV in a biological sample. Also provided are capture oligonucleotides
CC   {Seq ID. No. 10}-{Seq ID. No. 14} which are used in a a method for
CC   detecting HAV infection in a biological sample. The present sequence
CC   represents a detection probe specific for an internal control sequence.
XX
SQ   Sequence 21 BP; 5 A; 5 C; 6 G; 5 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 21;  DB 12;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 2.5;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              |||||||||||||||||||||
Db          1 CAGTGACATGCAGGTCTAGCT 21


RESULT 6
ADI53797
ID   ADI53797 standard; DNA; 23 BP.
XX
AC   ADI53797;
XX
DT   06-MAY-2004  (first entry)
XX
DE   HAV internal control specific detection probe.
XX
KW   HAV; nucleic acid amplification; nucleic acid detection; ss; probe.
XX
OS   Hepatitis A virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   modified_base   1
FT                   /*tag=  a
FT                   /mod_base= 5'-TET labelled
FT   modified_base   23
FT                   /*tag=  b
FT                   /mod_base= 3'-TAMRA labelled
XX
PN   WO2003106641-A2.
XX
PD   24-DEC-2003.
XX
PF   12-JUN-2003; 2003WO-US018827.
XX
PR   12-JUN-2002; 2002US-0388544P.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Shyamala V;
XX
DR   WPI; 2004-082181/08.
XX
PT   Hepatitis A virus specific primers and probes derived from conserved
PT   regions of the hepatitis A virus genome, useful in nucleic acid-based
PT   diagnostic tests for the detection of Hepatitis A virus in biological
PT   samples.
XX
PS   Example 2; SEQ ID NO 19; 44pp; English.
XX
CC   The invention relates to Hepatitis A virus (HAV) specific primers and
CC   probes derived from conserved regions of the hepatitis A virus genome.
CC   The HAV-specific primers and probes are used in a method for detecting
CC   HAV in a biological sample. Also provided are capture oligonucleotides
CC   {Seq ID. No. 10}-{Seq ID. No. 14} which are used in a a method for
CC   detecting HAV infection in a biological sample. The present sequence
CC   represents a detection probe specific for an internal control sequence.
XX
SQ   Sequence 23 BP; 5 A; 7 C; 6 G; 5 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 21;  DB 12;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 2.5;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              |||||||||||||||||||||
Db          3 CAGTGACATGCAGGTCTAGCT 23


RESULT 7
ADQ30670
ID   ADQ30670 standard; DNA; 23 BP.
XX
AC   ADQ30670;
XX
DT   23-SEP-2004  (first entry)
XX
DE   West Nile Virus internal control probe #1.
XX
KW   ss; primer; West Nile Virus; diagnosis.
XX
OS   West Nile virus.
XX
PN   WO2004055159-A2.
XX
PD   01-JUL-2004.
XX
PF   05-DEC-2003; 2003WO-US038750.
XX
PR   12-DEC-2002; 2002US-0432850P.
PR   20-JUN-2003; 2003US-0480431P.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Shyamala V;
XX
DR   WPI; 2004-488058/46.
XX
PT   New isolated oligonucleotides for accurately diagnosing West Nile virus
PT   infection or for capturing, detecting and quantitating West Nile virus in
PT   blood samples.
XX
PS   Claim 29; SEQ ID NO 40; 56pp; English.
XX
CC   The invention relates to an isolated oligonucleotide not more than 60
CC   nucleotides in length comprising a nucleotide sequence (S1) of at least
```

```
CC   10 contiguous nucleotides from any of the 28 nucleotide sequences (e.g.
CC   20, 21 or 23 bp) given in the specification derived from the West Nile
CC   Virus (WNV) genome, a nucleotide sequence (S2) having 90% sequence
CC   identity to the nucleotide sequence of (S1), or complements of (S1) and
CC   (S2). The oligonucleotide further comprises a detectable label at the 5'-
CC   end and/or the 3'-end. The detectable label is a fluorescent label
CC   selected from 6-carboxyfluorescein (6-FAM), tetramethyl rhodamine
CC   (TAMRA), and 2',4',5',7',-tetrachloro-4-7-dichlorofluorescein (TET). The
CC   composition and methods are useful for accurately diagnosing West Nile
CC   virus infection or for capturing, detecting and quantitating West Nile
CC   virus in biological samples, particularly blood samples. This sequence
CC   corresponds to a probe to the internal control sequence for the detection
CC   of WNV sequences using the oligonucleotides of the invention.
XX
SQ   Sequence 23 BP; 5 A; 5 C; 6 G; 5 T; 0 U; 2 Other;

  Query Match            100.0%;  Score 21;  DB 12;  Length 23;
  Best Local Similarity  100.0%;  Pred. No. 2.5;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              |||||||||||||||||||||
Db          2 CAGTGACATGCAGGTCTAGCT 22


RESULT 8
ADQ30671
ID   ADQ30671 standard; DNA; 25 BP.
XX
AC   ADQ30671;
XX
DT   23-SEP-2004  (first entry)
XX
DE   West Nile Virus internal control probe #2.
XX
KW   ss; primer; West Nile Virus; diagnosis.
XX
OS   West Nile virus.
XX
PN   WO2004055159-A2.
XX
PD   01-JUL-2004.
XX
PF   05-DEC-2003; 2003WO-US038750.
XX
PR   12-DEC-2002; 2002US-0432850P.
PR   20-JUN-2003; 2003US-0480431P.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Shyamala V;
XX
DR   WPI; 2004-488058/46.
XX
PT   New isolated oligonucleotides for accurately diagnosing West Nile virus
PT   infection or for capturing, detecting and quantitating West Nile virus in
PT   blood samples.
XX
PS   Claim 29; SEQ ID NO 41; 56pp; English.
XX
CC   The invention relates to an isolated oligonucleotide not more than 60
CC   nucleotides in length comprising a nucleotide sequence (S1) of at least
CC   10 contiguous nucleotides from any of the 28 nucleotide sequences (e.g.
CC   20, 21 or 23 bp) given in the specification derived from the West Nile
CC   Virus (WNV) genome, a nucleotide sequence (S2) having 90% sequence
CC   identity to the nucleotide sequence of (S1), or complements of (S1) and
CC   (S2). The oligonucleotide further comprises a detectable label at the 5'-
CC   end and/or the 3'-end. The detectable label is a fluorescent label
CC   selected from 6-carboxyfluorescein (6-FAM), tetramethyl rhodamine
CC   (TAMRA), and 2',4',5',7',-tetrachloro-4-7-dichlorofluorescein (TET). The
CC   composition and methods are useful for accurately diagnosing West Nile
CC   virus infection or for capturing, detecting and quantitating West Nile
CC   virus in biological samples, particularly blood samples. This sequence
CC   corresponds to a probe to the internal control sequence for the detection
CC   of WNV sequences using the oligonucleotides of the invention.
XX
SQ   Sequence 25 BP; 5 A; 7 C; 6 G; 5 T; 0 U; 2 Other;

  Query Match            100.0%;  Score 21;  DB 12;  Length 25;
  Best Local Similarity  100.0%;  Pred. No. 2.5;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              |||||||||||||||||||||
Db          4 CAGTGACATGCAGGTCTAGCT 24


RESULT 9
ABZ59634
ID   ABZ59634 standard; DNA; 681 BP.
XX
AC   ABZ59634;
XX
DT   22-APR-2003  (first entry)
XX
DE   Exemplary internal control nucleotide sequence SEQ ID NO:92.
XX
KW   Human parvovirus B19; parvovirus B19; infection; virus; blood; plasma;
KW   gene; ds.
XX
OS   Synthetic.
XX
PN   WO2003002753-A2.
XX
PD   09-JAN-2003.
XX
PF   28-JUN-2002; 2002WO-US020684.
XX
PR   28-JUN-2001; 2001US-0302077P.
PR   19-MAR-2002; 2002US-0365956P.
PR   29-MAR-2002; 2002US-0369224P.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Pichuantes S,  Shyamala V;
XX
DR   WPI; 2003-201510/19.
XX
PT   Detecting a human parvovirus B19 infection in a biological sample to
PT   prevent viral transmission, comprises reacting a parvovirus B19 nucleic
PT   acid with a primer complementary to the 3'-terminal portion of the RNA
PT   target sequence.
XX
PS   Claim 7; Fig 12; 148pp; English.
XX
CC   The present invention describes a method for detecting a human parvovirus
CC   B19 infection in a biological sample. The method comprises reacting the
CC   isolated parvovirus B19 nucleic acid with a first oligonucleotide
CC   consisting of a first primer containing a complexing sequence
CC   sufficiently complementary to the 3'-terminal portion of the RNA target
CC   sequence to complex with. Also described: (1) amplifying a target
CC   parvovirus B19 nucleotide sequence; (2) a polynucleotide comprising one
CC   of 47 700 base pair sequences (see ABZ59549 to ABZ59569, and ABZ59604 to
CC   ABZ59629); (3) a polynucleotide comprising either of 2 4678 base pair
CC   sequences (see ABZ59570 and ABZ59571); (4) an oligonucleotide primer
CC   consisting of a promoter region recognised by a DNA-dependent RNA
CC   polymerase operably linked to a human parvovirus B19-specific complexing
CC   sequence of 10-75 nucleotides; (5) an oligonucleotide probe comprising a
CC   parvovirus B19-specific hybridising sequence of 10-50 nucleotides linked
CC   to an acridinium ester label; and (6) a diagnostic test kit comprising an
CC   oligonucleotide primer of (4), and instructions for conducting the
CC   diagnostic test. The method is useful for detecting parvovirus infection
CC   in a biological sample, such as in blood products, to prevent
CC   transmission of the virus through blood and plasma derivatives or by
CC   close personal contact. ABZ59549 to ABZ59634 and ABP57262 to ABP57267
CC   represent sequences used in the exemplification of the present invention
```

```
XX
SQ   Sequence 681 BP; 206 A; 138 C; 137 G; 200 T; 0 U; 0 Other;

  Query Match            100.0%;  Score 21;  DB 9;  Length 681;
  Best Local Similarity  100.0%;  Pred. No. 3.4;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              |||||||||||||||||||||
Db         62 CAGTGACATGCAGGTCTAGCT 82


RESULT 10
ADI53795
ID   ADI53795 standard; DNA; 727 BP.
XX
AC   ADI53795;
XX
DT   06-MAY-2004  (first entry)
XX
DE   HAV internal control sequence.
XX
KW   HAV; nucleic acid amplification; nucleic acid detection; ds.
XX
OS   Hepatitis A virus.
OS   Synthetic.
XX
PN   WO2003106641-A2.
XX
PD   24-DEC-2003.
XX
PF   12-JUN-2003; 2003WO-US018827.
XX
PR   12-JUN-2002; 2002US-0388544P.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Shyamala V;
XX
DR   WPI; 2004-082181/08.
XX
PT   Hepatitis A virus specific primers and probes derived from conserved
PT   regions of the hepatitis A virus genome, useful in nucleic acid-based
PT   diagnostic tests for the detection of Hepatitis A virus in biological
PT   samples.
XX
PS   Example 2; SEQ ID NO 17; 44pp; English.
XX
CC   The invention relates to Hepatitis A virus (HAV) specific primers and
CC   probes derived from conserved regions of the hepatitis A virus genome.
CC   The HAV-specific primers and probes are used in a method for detecting
CC   HAV in a biological sample. Also provided are capture oligonucleotides
CC   {Seq ID. No. 10}-{Seq ID. No. 14} which are used in a a method for
CC   detecting HAV infection in a biological sample. The present sequence
CC   represents an internal control sequence used as a control for target
CC   capture and amplification.
XX
SQ   Sequence 727 BP; 147 A; 169 C; 186 G; 225 T; 0 U; 0 Other;

  Query Match            100.0%;  Score 21;  DB 12;  Length 727;
  Best Local Similarity  100.0%;  Pred. No. 3.4;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              |||||||||||||||||||||
Db        581 CAGTGACATGCAGGTCTAGCT 601


RESULT 11
ADQ30647
ID   ADQ30647 standard; DNA; 967 BP.
XX
AC   ADQ30647;
XX
DT   23-SEP-2004  (first entry)
XX
DE   West Nile virus internal diagnosis control sequence.
XX
KW   ss; internal control; West Nile Virus; diagnosis.
XX
OS   West Nile virus.
XX
PN   WO2004055159-A2.
XX
PD   01-JUL-2004.
XX
PF   05-DEC-2003; 2003WO-US038750.
XX
PR   12-DEC-2002; 2002US-0432850P.
PR   20-JUN-2003; 2003US-0480431P.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Shyamala V;
XX
DR   WPI; 2004-488058/46.
XX
PT   New isolated oligonucleotides for accurately diagnosing West Nile virus
PT   infection or for capturing, detecting and quantitating West Nile virus in
PT   blood samples.
XX
PS   Claim 27; SEQ ID NO 17; 56pp; English.
XX
CC   The invention relates to an isolated oligonucleotide not more than 60
CC   nucleotides in length comprising a nucleotide sequence (S1) of at least
CC   10 contiguous nucleotides from any of the 28 nucleotide sequences (e.g.
CC   20, 21 or 23 bp) given in the specification derived from the West Nile
CC   Virus (WNV) genome, a nucleotide sequence (S2) having 90% sequence
CC   identity to the nucleotide sequence of (S1), or complements of (S1) and
CC   (S2). The oligonucleotide further comprises a detectable label at the 5'-
CC   end and/or the 3'-end. The detectable label is a fluorescent label
CC   selected from 6-carboxyfluorescein (6-FAM), tetramethyl rhodamine
CC   (TAMRA), and 2',4',5',7',-tetrachloro-4-7-dichlorofluorescein (TET). The
CC   composition and methods are useful for accurately diagnosing West Nile
CC   virus infection or for capturing, detecting and quantitating West Nile
CC   virus in biological samples, particularly blood samples. This sequence
CC   corresponds to an internal control sequence for the detection of WNV
CC   sequences using the oligonucleotides of the invention.
XX
SQ   Sequence 967 BP; 273 A; 206 C; 272 G; 216 T; 0 U; 0 Other;

  Query Match            100.0%;  Score 21;  DB 12;  Length 967;
  Best Local Similarity  100.0%;  Pred. No. 3.5;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              |||||||||||||||||||||
Db        153 CAGTGACATGCAGGTCTAGCT 173


RESULT 12
ACC80539
ID   ACC80539 standard; DNA; 1696 BP.
XX
AC   ACC80539;
XX
DT   29-AUG-2003  (first entry)
XX
DE   Internal control region for hepatitis B virus DNA detection method.
XX
KW   Hepatitis B virus; diagnosis; nuclease assay; internal control region;
KW   transcription-mediated amplification; ds.
XX
OS   Hepatitis B virus.
XX
PN   WO2003031934-A2.
```

```
XX
PD   17-APR-2003.
XX
PF   09-OCT-2002; 2002WO-US032367.
XX
PR   09-OCT-2001; 2001US-0328492P.
PR   29-MAR-2002; 2002US-0368823P.
PR   02-JUL-2002; 2002US-0393561P.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Shyamala V;
XX
DR   WPI; 2003-403124/38.
XX
PT   New isolated hepatitis B virus (HBV) capture oligonucleotides, useful for
PT   detecting HBV infection in a biological sample, or in capturing HBV
PT   nucleic acids.
XX
PS   Disclosure; Fig 3; 50pp; English.
XX
CC   The invention relates to a novel method of detecting hepatitis B virus
CC   (HBV) infections in e.g. blood samples from donors, by capturing and
CC   amplifying conserved regions of the HBV genome using a transcription-
CC   mediated amplification (TMA) method as well as a 5' nuclease assay. The
CC   method may also use an internal control sequence (this sequence), to
CC   determine the level of amplification and detection by the primers and
CC   probes used in the method of the invention. The new method is very
CC   sensitive and is able to detect about 100 infectious units (IU) of HBV in
CC   a viremic sample
XX
SQ   Sequence 1696 BP; 359 A; 462 C; 386 G; 489 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 21;  DB 8;  Length 1696;
  Best Local Similarity   100.0%;  Pred. No. 3.7;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAGCT 21
              |||||||||||||||||||||
Db       1428 CAGTGACATGCAGGTCTAGCT 1448


RESULT 13
ADE71204/c
ID   ADE71204 standard; DNA; 5446 BP.
XX
AC   ADE71204;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Novel human protein coding sequence #20.
XX
KW   human; novel protein; drug; gene; ds.
XX
OS   Homo sapiens.
XX
PN   JP2002345493-A.
XX
PD   03-DEC-2002.
XX
PF   29-MAR-2001; 2002JP-00049046.
XX
PR   29-MAR-2001; 2001JP-00095524.
XX
PA   (KAZU-) ZH KAZUSA DNA KENKYUSHO.
XX
DR   WPI; 2003-460885/44.
DR   P-PSDB; ADE71266.
XX
PT   A gene and a protein encoded by it, used in drugs.
XX
PS   Claim 1; SEQ ID NO 20; 257pp; Japanese.
XX
CC   The invention comprises the amino acid and coding sequences of novel
CC   human proteins. The DNA and protein sequences of the invention are used
CC   in drugs. The present DNA sequence encodes a novel human protein of the
CC   invention.
XX
SQ   Sequence 5446 BP; 1477 A; 1320 C; 1292 G; 1357 T; 0 U; 0 Other;

  Query Match             82.9%;  Score 17.4;  DB 10;  Length 5446;
  Best Local Similarity   94.7%;  Pred. No. 2.3e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAG 19
              |||||||| ||||||||||
Db        833 CAGTGACAGGCAGGTCTAG 815


RESULT 14
ABZ73855/c
ID   ABZ73855 standard; DNA; 28482 BP.
XX
AC   ABZ73855;
XX
DT   12-MAY-2003  (first entry)
XX
DE   Secreted protein gene 64 genomic fragment HCEGX05, SEQ ID NO:1002.
XX
KW   Human; secreted protein; cancer; tumour; hyperproliferative disorder;
KW   autoimmune disorder; inflammation; angiogenic diseases; AIDS;
KW   acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
KW   drug screening; chromosome identification; chromosome mapping;
KW   cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;
KW   antianaemic; vulnerary; chromosome 20q13; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200277013-A2.
XX
PD   03-OCT-2002.
XX
PF   26-MAR-2002; 2002WO-US009370.
XX
PR   27-MAR-2001; 2001US-0278650P.
PR   12-SEP-2001; 2001US-00950082.
PR   12-SEP-2001; 2001US-00950083.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2003-040578/03.
XX
PT   New human secreted proteins and nucleic acids, useful for detecting or
PT   treating cancer or other hyperproliferative disorders, autoimmune
PT   disorders, inflammatory disorders, HIV disease, hepatitis or anemia.
XX
PS   Disclosure; Page 1651-1658; 2474pp; English.
XX
CC   ABZ73281-ABZ73697 represent cDNAs corresponding to 391 human secreted
CC   protein genes, and ABP00947-ABP01363 represent the proteins they encode.
CC   ABZ73698-ABZ74687 represent human secreted protein genomic fragments. The
CC   invention also encompasses antibodies specific for the secreted proteins,
CC   the use of the secreted proteins in drug screening and recombinant
CC   vectors and host cells comprising a nucleic acid of the invention. The
CC   secreted proteins are thought to be involved in biological activities
CC   associated with cellular signalling, cellular differentiation, cell
CC   migration, prohormone activation and neurotransmitter activity. The
CC   secreted proteins, nucleic acids encoding them, antibodies or antibody
CC   fragments specific for the secreted proteins, and modulators of protein
CC   activity are useful for diagnosing or treating cancers or other
CC   hyperproliferative disorders. Additionally, the secreted proteins and
CC   their nucleic acids may also be used in the treatment of autoimmune
CC   disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
CC   (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
```

```
CC   wound healing. Nucleic acids of the invention may be used for chromosome
CC   identification, chromosome mapping, in gene therapy, for identifying
CC   individuals from minute biological samples, as hybridisation probes, and
CC   as molecular weight markers. The present sequence represents a human
CC   secreted protein genomic fragment referred to in the disclosure of the
CC   invention
XX
SQ   Sequence 28482 BP; 7636 A; 6245 C; 6763 G; 7838 T; 0 U; 0 Other;

  Query Match             82.9%;  Score 17.4;  DB 8;  Length 28482;
  Best Local Similarity   94.7%;  Pred. No. 2.6e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAG 19
              ||||||||| ||||||||||
Db      12331 CAGTGACAGGCAGGTCTAG 12313


RESULT 15
ADA44262/c
ID   ADA44262 standard; DNA; 28482 BP.
XX
AC   ADA44262;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human secreted protein DNA SEQ ID 455.
XX
KW   Gene therapy; human; Antidiabetic; Anorectic; Ophthalmological;
KW   Neuroprotective; Cerebroprotective; Antianemic; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003000865-A2.
XX
PD   03-JAN-2003.
XX
PF   26-MAR-2002; 2002WO-US009105.
XX
PR   27-MAR-2001; 2001US-0278650P.
PR   12-SEP-2001; 2001US-00950082.
PR   12-SEP-2001; 2001US-00950083.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2003-184045/18.
XX
PT   A human secreted protein and nucleic acids useful for preparing a
PT   diagnostic or pharmaceutical composition for diagnosing or treating
PT   diabetes or conditions related to diabetes, e.g. hyperglycemia, obesity,
PT   retinopathy, neuropathy.
XX
PS   Disclosure; SEQ ID NO 455; 701pp; English.
XX
CC   The invention relates to novel genes and their fragments which are useful
CC   for preventing, treating or ameliorating medical conditions e.g. by
CC   protein or gene therapy. The genes are isolated from a range of human
CC   tissues disclosed in the specification. The nucleic acids and proteins
CC   are useful in the diagnosis, treatment and prevention of conditions
CC   related to diabetes, e.g. hyperglycaemia, obesity, retinopathy,
CC   polyneuropathy, atherosclerosis, anaemia, stroke, gangrene, impotence,
CC   infection, cataract, renal disorders, or endocrine disorders. The present
CC   sequence was used to illustrate the invention.
XX
SQ   Sequence 28482 BP; 7636 A; 6245 C; 6763 G; 7838 T; 0 U; 0 Other;

  Query Match             82.9%;  Score 17.4;  DB 8;  Length 28482;
  Best Local Similarity   94.7%;  Pred. No. 2.6e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAG 19
              ||||||||| ||||||||||
Db      12331 CAGTGACAGGCAGGTCTAG 12313


RESULT 16
ABZ74517/c
ID   ABZ74517 standard; DNA; 32681 BP.
XX
AC   ABZ74517;
XX
DT   12-MAY-2003  (first entry)
XX
DE   Secreted protein gene 329 genomic fragment HTLBT80, SEQ ID NO:1664.
XX
KW   Human; secreted protein; cancer; tumour; hyperproliferative disorder;
KW   autoimmune disorder; inflammation; angiogenic diseases; AIDS;
KW   acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
KW   drug screening; chromosome identification; chromosome mapping;
KW   cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;
KW   antianaemic; vulnerary; chromosome 20q11.21-13.11; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200277013-A2.
XX
PD   03-OCT-2002.
XX
PF   26-MAR-2002; 2002WO-US009370.
XX
PR   27-MAR-2001; 2001US-0278650P.
PR   12-SEP-2001; 2001US-00950082.
PR   12-SEP-2001; 2001US-00950083.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2003-040578/03.
XX
PT   New human secreted proteins and nucleic acids, useful for detecting or
PT   treating cancer or other hyperproliferative disorders, autoimmune
PT   disorders, inflammatory disorders, HIV disease, hepatitis or anemia.
XX
PS   Disclosure; Page 2245-2253; 2474pp; English.
XX
CC   ABZ73281-ABZ73697 represent cDNAs corresponding to 391 human secreted
CC   protein genes, and ABP00947-ABP01363 represent the proteins they encode.
CC   ABZ73698-ABZ74687 represent human secreted protein genomic fragments. The
CC   invention also encompasses antibodies specific for the secreted proteins,
CC   the use of the secreted proteins in drug screening and recombinant
CC   vectors and host cells comprising a nucleic acid of the invention. The
CC   secreted proteins are thought to be involved in biological activities
CC   associated with cellular signalling, cellular differentiation, cell
CC   migration, prohormone activation and neurotransmitter activity. The
CC   secreted proteins, nucleic acids encoding them, antibodies or antibody
CC   fragments specific for the secreted proteins, and modulators of protein
CC   activity are useful for diagnosing or treating cancers or other
CC   hyperproliferative disorders. Additionally, the secreted proteins and
CC   their nucleic acids may also be used in the treatment of autoimmune
CC   disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
CC   (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
CC   wound healing. Nucleic acids of the invention may be used for chromosome
CC   identification, chromosome mapping, in gene therapy, for identifying
CC   individuals from minute biological samples, as hybridisation probes, and
CC   as molecular weight markers. The present sequence represents a human
CC   secreted protein genomic fragment referred to in the disclosure of the
CC   invention
XX
SQ   Sequence 32681 BP; 8783 A; 7103 C; 7721 G; 9074 T; 0 U; 0 Other;

  Query Match             82.9%;  Score 17.4;  DB 8;  Length 32681;
  Best Local Similarity   94.7%;  Pred. No. 2.7e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy          1 CAGTGACATGCAGGTCTAG 19
              ||||||||| ||||||||||
Db      16526 CAGTGACAGGCAGGTCTAG 16508


RESULT 17
ABZ73854/c
ID   ABZ73854 standard; DNA; 32681 BP.
XX
AC   ABZ73854;
XX
DT   12-MAY-2003  (first entry)
XX
DE   Secreted protein gene 64 genomic fragment HCEGX05, SEQ ID NO:1001.
XX
KW   Human; secreted protein; cancer; tumour; hyperproliferative disorder;
KW   autoimmune disorder; inflammation; angiogenic diseases; AIDS;
KW   acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
KW   drug screening; chromosome identification; chromosome mapping;
KW   cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;
KW   antianaemic; vulnerary; chromosome 20q13; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200277013-A2.
XX
PD   03-OCT-2002.
XX
PF   26-MAR-2002; 2002WO-US009370.
XX
PR   27-MAR-2001; 2001US-0278650P.
PR   12-SEP-2001; 2001US-00950082.
PR   12-SEP-2001; 2001US-00950083.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2003-040578/03.
XX
PT   New human secreted proteins and nucleic acids, useful for detecting or
PT   treating cancer or other hyperproliferative disorders, autoimmune
PT   disorders, inflammatory disorders, HIV disease, hepatitis or anemia.
XX
PS   Disclosure; Page 1643-1651; 2474pp; English.
XX
CC   ABZ73281-ABZ73697 represent cDNAs corresponding to 391 human secreted
CC   protein genes, and ABP00947-ABP01363 represent the proteins they encode.
CC   ABZ73698-ABZ74687 represent human secreted protein genomic fragments. The
CC   invention also encompasses antibodies specific for the secreted proteins,
CC   the use of the secreted proteins in drug screening and recombinant
CC   vectors and host cells comprising a nucleic acid of the invention. The
CC   secreted proteins are thought to be involved in biological activities
CC   associated with cellular signalling, cellular differentiation, cell
CC   migration, prohormone activation and neurotransmitter activity. The
CC   secreted proteins, nucleic acids encoding them, antibodies or antibody
CC   fragments specific for the secreted proteins, and modulators of protein
CC   activity are useful for diagnosing or treating cancers or other
CC   hyperproliferative disorders. Additionally, the secreted proteins and
CC   their nucleic acids may also be used in the treatment of autoimmune
CC   disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
CC   (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
CC   wound healing. Nucleic acids of the invention may be used for chromosome
CC   identification, chromosome mapping, in gene therapy, for identifying
CC   individuals from minute biological samples, as hybridisation probes, and
CC   as molecular weight markers. The present sequence represents a human
CC   secreted protein genomic fragment referred to in the disclosure of the
CC   invention
XX
SQ   Sequence 32681 BP; 8783 A; 7103 C; 7721 G; 9074 T; 0 U; 0 Other;

  Query Match             82.9%;  Score 17.4;  DB 8;  Length 32681;
  Best Local Similarity   94.7%;  Pred. No. 2.7e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAG 19
              ||||||||| ||||||||||
Db      16526 CAGTGACAGGCAGGTCTAG 16508


RESULT 18
ADA98915/c
ID   ADA98915 standard; DNA; 32681 BP.
XX
AC   ADA98915;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human secreted protein-related DNA sequence #508.
XX
KW   human; secreted protein; cardiovascular disorder; arrhythmia;
KW   atherosclerosis; stroke; endocarditis; congestive heart failure;
KW   rheumatic heart disease; cardiomyopathy; hemorrhoids; varicose veins;
KW   migraine; thrombosis; neural disorder; immune system disorder;
KW   muscular disorder; reproductive disorder; gastrointestinal disorder;
KW   pulmonary disorder; renal disorder; proliferative disorder; cancer; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003004623-A2.
XX
PD   16-JAN-2003.
XX
PF   26-MAR-2002; 2002WO-US009922.
XX
PR   27-MAR-2001; 2001US-0278650P.
PR   12-SEP-2001; 2001US-00950082.
PR   12-SEP-2001; 2001US-00950083.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2003-247946/24.
XX
PT   New human secreted polypeptide and nucleic acid molecules, useful for
PT   diagnosing, preventing, prognosticating or treating cardiovascular
PT   disorders (e.g. arrhythmia, atherosclerosis, cardiomyopathy, or
PT   thrombosis).
XX
PS   Disclosure; SEQ ID NO 1024; 1572pp; English.
XX
CC   The invention comprises the amino acid and coding sequence of human
CC   secreted proteins. The DNA and protein sequences of the invention are
CC   useful in the treatment of cardiovascular disorders, such as: arrhythmia,
CC   atherosclerosis, stroke, endocarditis, congestive heart failure,
CC   rheumatic heart disease, cardiomyopathy, hemorrhoids, varicose veins,
CC   migraine, or thrombosis. The DNA and protein sequences may also be used
CC   for treating or preventing: neural disorders, immune system disorders,
CC   muscular disorders, reproductive disorders, gastrointestinal disorders,
CC   pulmonary disorders, renal disorders, proliferative disorders and/or
CC   cancerous diseases. The present DNA sequence is used in the
CC   exemplification of the invention. NOTE: The present sequence is shown on
CC   the WIPO website.
XX
SQ   Sequence 32681 BP; 8783 A; 7103 C; 7721 G; 9074 T; 0 U; 0 Other;

  Query Match             82.9%;  Score 17.4;  DB 8;  Length 32681;
  Best Local Similarity   94.7%;  Pred. No. 2.7e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAG 19
              ||||||||| ||||||||||
Db      16526 CAGTGACAGGCAGGTCTAG 16508
```

```
RESULT 19
ADA44261/c
ID   ADA44261 standard; DNA; 32681 BP.
XX
AC   ADA44261;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human secreted protein DNA SEQ ID 454.
XX
KW   Gene therapy; human; Antidiabetic; Anorectic; Ophthalmological;
KW   Neuroprotective; Cerebroprotective; Antianemic; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003000865-A2.
XX
PD   03-JAN-2003.
XX
PF   26-MAR-2002; 2002WO-US009105.
XX
PR   27-MAR-2001; 2001US-0278650P.
PR   12-SEP-2001; 2001US-00950082.
PR   12-SEP-2001; 2001US-00950083.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2003-184045/18.
XX
PT   A human secreted protein and nucleic acids useful for preparing a
PT   diagnostic or pharmaceutical composition for diagnosing or treating
PT   diabetes or conditions related to diabetes, e.g. hyperglycemia, obesity,
PT   retinopathy, neuropathy.
XX
PS   Disclosure; SEQ ID NO 454; 701pp; English.
XX
CC   The invention relates to novel genes and their fragments which are useful
CC   for preventing, treating or ameliorating medical conditions e.g. by
CC   protein or gene therapy. The genes are isolated from a range of human
CC   tissues disclosed in the specification. The nucleic acids and proteins
CC   are useful in the diagnosis, treatment and prevention of conditions
CC   related to diabetes, e.g. hyperglycaemia, obesity, retinopathy,
CC   polyneuropathy, atherosclerosis, anaemia, stroke, gangrene, impotence,
CC   infection, cataract, renal disorders, or endocrine disorders. The present
CC   sequence was used to illustrate the invention.
XX
SQ   Sequence 32681 BP; 8783 A; 7103 C; 7721 G; 9074 T; 0 U; 0 Other;

  Query Match             82.9%;  Score 17.4;  DB 8;  Length 32681;
  Best Local Similarity   94.7%;  Pred. No. 2.7e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAG 19
              ||||||||| ||||||||||
Db      16526 CAGTGACAGGCAGGTCTAG 16508


RESULT 20
ADA44519/c
ID   ADA44519 standard; DNA; 32681 BP.
XX
AC   ADA44519;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human secreted protein DNA SEQ ID 712.
XX
KW   Gene therapy; human; Antidiabetic; Anorectic; Ophthalmological;
KW   Neuroprotective; Cerebroprotective; Antianemic; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003000865-A2.
XX
PD   03-JAN-2003.
XX
PF   26-MAR-2002; 2002WO-US009105.
XX
PR   27-MAR-2001; 2001US-0278650P.
PR   12-SEP-2001; 2001US-00950082.
PR   12-SEP-2001; 2001US-00950083.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2003-184045/18.
XX
PT   A human secreted protein and nucleic acids useful for preparing a
PT   diagnostic or pharmaceutical composition for diagnosing or treating
PT   diabetes or conditions related to diabetes, e.g. hyperglycemia, obesity,
PT   retinopathy, neuropathy.
XX
PS   Disclosure; SEQ ID NO 712; 701pp; English.
XX
CC   The invention relates to novel genes and their fragments which are useful
CC   for preventing, treating or ameliorating medical conditions e.g. by
CC   protein or gene therapy. The genes are isolated from a range of human
CC   tissues disclosed in the specification. The nucleic acids and proteins
CC   are useful in the diagnosis, treatment and prevention of conditions
CC   related to diabetes, e.g. hyperglycaemia, obesity, retinopathy,
CC   polyneuropathy, atherosclerosis, anaemia, stroke, gangrene, impotence,
CC   infection, cataract, renal disorders, or endocrine disorders. The present
CC   sequence was used to illustrate the invention.
XX
SQ   Sequence 32681 BP; 8783 A; 7103 C; 7721 G; 9074 T; 0 U; 0 Other;

  Query Match             82.9%;  Score 17.4;  DB 8;  Length 32681;
  Best Local Similarity   94.7%;  Pred. No. 2.7e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAGTGACATGCAGGTCTAG 19
              ||||||||| ||||||||||
Db      16526 CAGTGACAGGCAGGTCTAG 16508


RESULT 21
ADC20949/c
ID   ADC20949 standard; DNA; 32681 BP.
XX
AC   ADC20949;
XX
DT   18-DEC-2003  (first entry)
XX
DE   Human secreted protein-related DNA sequence #367.
XX
KW   gene therapy; human; secreted protein; haemopoietic disorder;
KW   haematological disorder; anaemia; haemophilia; inflammatory disorder;
KW   inflammatory bowel disease; Crohn's disease; neoplastic disease; cancer;
KW   leukaemia; wound healing; epithelial cell proliferation disorder;
KW   immune disorder; autoimmune disorder; asthmatic disorder;
KW   cardiovascular disorder; atherosclerosis; myocarditis;
KW   infectious disease; HIV; AIDS; endocrine disorder; diabetes;
KW   gastrointestinal disorder; duodenal ulcer; gastroenteritis; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200292787-A2.
XX
PD   21-NOV-2002.
XX
PF   26-MAR-2002; 2002WO-US009257.
XX
```

PR 27-MAR-2001; 2001US-0278650P.
PR 12-SEP-2001; 2001US-00950082.
PR 12-SEP-2001; 2001US-00950083.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-129287/12.
XX
PT New human secreted proteins and nucleic acid molecules, useful for
PT preparing a diagnostic or pharmaceutical composition for diagnosing,
PT preventing or treating hematopoietic or hematologic disorders, e.g.
PT anemia or hemophilia.
XX
PS Disclosure; SEQ ID NO 903; 1512pp; English.
XX
CC The invention comprises the amino acid and coding sequences of human
CC secreted proteins. The DNA and protein sequences of the invention are
CC useful for detecting, preventing, diagnosing, prognosticating, treating
CC or ameliorating: haematopoietic or haematological disorders (e.g. anaemia
CC and haemophilia); inflammatory disorders (e.g. inflammatory bowel disease
CC and Crohn's disease); neoplastic disease (e.g. cancer and leukaemia);
CC wound healing and disorders of epithelial cell proliferation; immune
CC disorders (e.g. autoimmune disorders and asthmatic disorders);
CC cardiovascular disorders (e.g. atherosclerosis and myocarditis);
CC infectious disease (e.g. HIV/AIDS); endocrine disorders (e.g. diabetes);
CC and gastrointestinal disorders (e.g. duodenal ulcers and
CC gastroenteritis). The present DNA sequence was used in the
CC exemplification of the invention.
XX
SQ Sequence 32681 BP; 8783 A; 7103 C; 7721 G; 9074 T; 0 U; 0 Other;

Query Match 82.9%; Score 17.4; DB 10; Length 32681;
Best Local Similarity 94.7%; Pred. No. 2.7e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAGTGACATGCAGGTCTAG 19
||||||| ||||||||||
Db 16526 CAGTGACAGGCAGGTCTAG 16508

RESULT 22
ABZ68053/c
ID ABZ68053 standard; DNA; 32681 BP.
XX
AC ABZ68053;
XX
DT 26-MAR-2003 (first entry)
XX
DE Human secreted protein encoding genomic DNA SEQ ID NO 1576.
XX
KW Human; secreted protein; nootropic; neuroprotective; cytostatic;
KW virucide; dermatological; immunosuppressive; antiinflammatory; anti-HIV;
KW vulnerary; antibacterial; antiparkinsonian; antisickling; antianaemic;
KW antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective;
KW antiinflammatory; antiallergic; antidiabetic; antiulcer; anticonvulsant;
KW antifungal; antiparasitic; cardiant; immune disorder; infection; vaccine;
KW cardiovascular disorder; neurological disease; nephrotropic;
KW gene therapy; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200277186-A2.
XX
PD 03-OCT-2002.
XX
PF 26-MAR-2002; 2002WO-US009188.
XX
PR 27-MAR-2001; 2001US-0278650P.
PR 12-SEP-2001; 2001US-00950082.
PR 12-SEP-2001; 2001US-00950083.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-040583/03.
XX
PT New human secreted proteins encoded by genes contained in cDNA clones
PT (e.g. HGCAC19), useful for preventing, treating or diagnosing e.g. AIDS,
PT multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or
PT West Nile fever.
XX
PS Disclosure; Page 2201-2209; 2423pp; English.
XX
CC The invention relates to novel human genes (ABZ66891-ABZ68209) and the
CC encoded secreted proteins (ABP99470-ABP99872) useful for preventing,
CC treating or ameliorating medical conditions e.g. by protein or gene
CC therapy. The genes are isolated from a range of human tissues disclosed
CC in the specification. The nucleic acids, proteins, antibodies and
CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
CC lung or urogenital; (b) immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
CC myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.
CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
CC bacterial, fungal and parasitic infections
XX
SQ Sequence 32681 BP; 8783 A; 7103 C; 7721 G; 9074 T; 0 U; 0 Other;

Query Match 82.9%; Score 17.4; DB 10; Length 32681;
Best Local Similarity 94.7%; Pred. No. 2.7e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAGTGACATGCAGGTCTAG 19
||||||| ||||||||||
Db 16526 CAGTGACAGGCAGGTCTAG 16508

RESULT 23
ABX46275
ID ABX46275 standard; cDNA; 320 BP.
XX
AC ABX46275;
XX
DT 21-FEB-2003 (first entry)
XX
DE Bovine EST associated with lactation/muscle/fat deposition #11440.
XX
KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW muscle deposition; fat deposition; genome mapping; gene identification;
KW gene analysis; cattle breeding.
XX
OS Bos Taurus.
XX
PN US2002137139-A1.
XX
PD 26-SEP-2002.
XX
PF 24-SEP-2001; 2001US-00960352.
XX
PR 12-JAN-1999; 99US-0115707P.
PR 11-JAN-2000; 2000US-00480902.
XX
PA (BYAT/) BYATT J C.
PA (MATH/) MATHIALAGAN N.
PA (TAON/) TAO N.
PA (WARR/) WARREN W C.
XX
PI Byatt JC, Mathialagan N, Tao N, Warren WC;
XX
DR WPI; 2003-110599/10.

```
XX
PT   New nucleic acid associated with lactation, and muscle and fat
PT   deposition, useful for genome mapping, gene identification and analysis,
PT   cattle breeding, or for genetically improving cattle.
XX
PS   Claim 2; SEQ ID NO 11440; 245pp; English.
XX
CC   The invention relates to a purified nucleic acid molecule associated with
CC   lactation or muscle and fat deposition (designated LMFD), derived from
CC   cattle, and the LMFD nucleic acid can specifically hybridise to a second
CC   nucleic acid molecule comprising any of 15112 nucleotide sequences,
CC   appearing as ABX34836-ABX49947, or complements of them. Also included are
CC   ; (1) a transformed cell having a nucleic acid comprising an LMFD nucleic
CC   acid linked to a promoter and a 3' non- translated sequence that
CC   functions in the cell to cause termination of transcription and addition
CC   of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
CC   (2) determining a level or pattern of a molecule in a bovine cell or
CC   tissue comprising: (a) incubating a marker nucleic acid (comprising any
CC   of the 15112 nucleic acid sequences or its complement or fragment) with a
CC   complementary nucleic acid molecule obtained from the bovine cell or
CC   tissue, where hybridisation between the marker nucleic acid and the
CC   complementary nucleic acid permits the detection of the molecule; and (b)
CC   detecting the level or pattern of the complementary nucleic acid, where
CC   the detection of the complementary nucleic acid is predictive of the
CC   level or pattern of the molecule. The LMFD nucleic acid is used for
CC   determining a level or pattern of a molecule in a bovine cell or tissue.
CC   It is useful for genome mapping, gene identification and analysis, cattle
CC   breeding, preparation of constructs for use in cattle gene expression, or
CC   for genetically improving cattle. The present sequence is one of the
CC   15112 bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The
CC   present sequence was not shown in the specification but was obtained in
CC   electronic format from the USPTO web site:
CC   seqdata.uspto.gov/sequence.html?DocID=20020137139
XX
SQ   Sequence 320 BP; 78 A; 80 C; 97 G; 65 T; 0 U; 0 Other;

  Query Match             80.0%;  Score 16.8;  DB 8;  Length 320;
  Best Local Similarity   90.0%;  Pred. No. 3.4e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 AGTGACATGCAGGTCTAGCT 21
              | ||||||||||||||| ||
Db        144 AATGACATGCAGGTCTACCT 163


RESULT 24
AAS00624
ID   AAS00624 standard; DNA; 36221 BP.
XX
AC   AAS00624;
XX
DT   07-SEP-2001  (first entry)
XX
DE   Human death-associated protein 6 (DAXX) gene.
XX
KW   Death-associated protein 6; DAXX; polymorphism; haplotype pair; human;
KW   immune disorder; autoimmune disease; population diversity; ds;
KW   paternity testing; anthropological lineage; forensic application.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(26869,G)
FT                   /*tag=  a
FT   variation       replace(26870,T)
FT                   /*tag=  b
FT   variation       replace(27145,A)
FT                   /*tag=  c
FT   variation       replace(27239,G)
FT                   /*tag=  d
FT   variation       replace(27620,T)
FT                   /*tag=  e
FT   variation       replace(27788,G)
FT                   /*tag=  f
FT   variation       replace(27806,T)
FT                   /*tag=  g
FT   variation       replace(27816,T)
FT                   /*tag=  h
FT   variation       replace(27869,T)
FT                   /*tag=  i
FT   variation       replace(27905,A)
FT                   /*tag=  j
FT   variation       replace(27916,C)
FT                   /*tag=  k
FT   variation       replace(28194,T)
FT                   /*tag=  l
FT   variation       replace(28339,T)
FT                   /*tag=  m
FT   variation       replace(28470,C)
FT                   /*tag=  n
FT   variation       replace(29010,T)
FT                   /*tag=  o
FT   variation       replace(30235,T)
FT                   /*tag=  p
FT   variation       replace(30665,A)
FT                   /*tag=  q
FT   variation       replace(30666,T)
FT                   /*tag=  r
FT   variation       replace(30752,T)
FT                   /*tag=  s
FT   variation       replace(31916,T)
FT                   /*tag=  t
XX
PN   WO200125245-A2.
XX
PD   12-APR-2001.
XX
PF   05-OCT-2000; 2000WO-US027487.
XX
PR   06-OCT-1999;   99US-0157909P.
XX
PA   (GENA-) GENAISSANCE PHARM INC.
XX
PI   Chew A,  Choi JY,  Denton RR,  Nandabalan K,  Stephens JC;
XX
DR   WPI; 2001-308220/32.
XX
PT   New human death-associated protein 6 (DAXX) gene variants comprising 19
PT   polymorphic sites useful in studying the effect of variation on the
PT   biological activity of DAXX and in developing drugs targeting the
PT   protein.
XX
PS   Claim 1; Fig 1; 97pp; English.
XX
CC   The sequence represents a DNA encoding human death-associated protein 6
CC   (DAXX). This gene may comprise one or more polymorphisms at specific
CC   nucleotide positions to form one of nineteen possible polymorphic
CC   variants. Associations between a trait and a genotype or a haplotype of
CC   the DAXX gene can be identified by comparing the frequency of the
CC   genotype or haplotype in a population exhibiting the trait with that of a
CC   reference population. A higher frequency in the trait population
CC   indicates an association. Methods involving genotyping or haplotyping of
CC   the DAXX gene of an individual can lead to prediction of haplotype pairs
CC   for the DAXX gene of related individuals, and may be useful in studying
CC   the expression and biological function of DAXX, as well as in developing
CC   drugs targeting this protein. Polymorphic variants of DAXX are useful in
CC   studying the effect of the variation on the biological activity of DAXX
CC   as well as on the binding affinity of candidate drugs targeting DAXX for
CC   the treatment of autoimmune diseases and other immune disorders.
CC   Polymorphism is also useful for studying population diversity,
CC   anthropological lineage, paternity testing, forensic applications, and
CC   for identifying associations between the DAXX genetic variation and a
CC   trait such as level of drug response or susceptibility to disease. DAXX
CC   proteins may be used to measure binding affinities of one or more
CC   candidate drugs targeting the DAXX protein
XX
```

```
SQ   Sequence 36221 BP; 8897 A; 8473 C; 9437 G; 9414 T; 0 U; 0 Other;

  Query Match            80.0%;  Score 16.8;  DB 4;  Length 36221;
  Best Local Similarity  90.0%;  Pred. No. 5.2e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 AGTGACATGCAGGTCTAGCT 21
              ||||||| |||| |||||||
Db      21390 AGTGACAGGCAGTTCTAGCT 21409


RESULT 25
ACN45158
ID   ACN45158 standard; DNA; 72705 BP.
XX
AC   ACN45158;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human genomic sequence hCG25130.
XX
KW   Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003073826-A2.
XX
PD   12-SEP-2003.
XX
PF   28-FEB-2003; 2003WO-US006235.
XX
PR   01-MAR-2002; 2002US-00087192.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-328604/31.
XX
PT   Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT   comprises a nucleotide sequence.
XX
PS   Claim 1; SEQ ID NO 1966; 0pp; English.
XX
CC   The present invention relates to novel DNA and protein sequences which
CC   are associated with carcinomas. The sequences are useful for: (i) for
CC   screening drug candidates; (ii) for screening of bioactive agent capable
CC   of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC   a bioactive agent capable of modulating the activity of CAP; (iv) for
CC   evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC   carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC   carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC   (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC   determining Carcinoma Associated (CA) gene copy number. In addition, the
CC   CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC   carcinoma including lymphoma. The present sequence is one such CA coding
CC   sequence. Note: This patent is an equivalent to basic patent
CC   US2002182586A1, for which no sequence data was published
XX
SQ   Sequence 72705 BP; 18277 A; 18952 C; 18052 G; 17424 T; 0 U; 0 Other;

  Query Match            78.1%;  Score 16.4;  DB 11;  Length 72705;
  Best Local Similarity  94.4%;  Pred. No. 8.7e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 AGTGACATGCAGGTCTAG 19
              ||| ||||||||||||||
Db      32065 AGTCACATGCAGGTCTAG 32082


Search completed: September  6, 2005, 20:39:33
Job time : 194.656 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 6, 2005, 17:45:55 ; Search time 1572.31 Seconds
(without alignments)
532.600 Million cell updates/sec

Title: US-10-729-421-35
Perfect score: 22
Sequence: 1 agccctcttcagtccaatcaag 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	19	86.4	1174	8	CC229626	CC229626 CH261-46H
	2	18.8	85.5	767	2	BE541043	BE541043 601064391
	3	18.8	85.5	872	2	BF687801	BF687801 602066853
c	4	18.4	83.6	935	9	CL901625	CL901625 CSHC1000
c	5	17.8	80.9	169	1	AA693379	AA693379 ah21g06.s
c	6	17.8	80.9	185	2	BF833890	BF833890 RC1-HT088
	7	17.8	80.9	193	2	AW800621	AW800621 MR1-UM006
	8	17.8	80.9	205	2	BF111499	BF111499 7l28f04.x
c	9	17.8	80.9	270	1	AA884812	AA884812 am28b05.s
c	10	17.8	80.9	270	2	AW819544	AW819544 RC5-ST029
c	11	17.8	80.9	271	1	AA437133	AA437133 zv53e10.s
	12	17.8	80.9	271	2	AW467041	AW467041 ha08c03.x
	13	17.8	80.9	289	2	BB720494	BB720494 BB720494
	14	17.8	80.9	295	4	BM151302	BM151302 TCBAP1D61
c	15	17.8	80.9	309	1	AI203209	AI203209 qr23h08.x
	16	17.8	80.9	310	7	T46996	T46996 yb12b06.s1
	17	17.8	80.9	328	1	AA969466	AA969466 oo81d01.s
c	18	17.8	80.9	328	2	AW615159	AW615159 hg73h02.x
	19	17.8	80.9	339	1	AI307235	AI307235 tb18c07.x
c	20	17.8	80.9	339	1	AI631503	AI631503 wa89g09.x
c	21	17.8	80.9	339	1	AI633750	AI633750 tt28b04.x
c	22	17.8	80.9	339	1	AI954798	AI954798 wq33d07.x
c	23	17.8	80.9	339	1	AI954834	AI954834 wq33h04.x
c	24	17.8	80.9	339	1	AI962281	AI962281 wq46e04.x
c	25	17.8	80.9	339	7	T46995	T46995 yb12b06.r1
c	26	17.8	80.9	340	2	AW291932	AW291932 UI-H-BI2-
	27	17.8	80.9	345	2	AW237469	AW237469 xm72c11.x
c	28	17.8	80.9	350	8	AZ693999	AZ693999 AST-1HDB2
	29	17.8	80.9	356	1	AA749323	AA749323 ny12c01.s
c	30	17.8	80.9	376	1	AI150471	AI150471 qf41d04.x
	31	17.8	80.9	389	1	AA768437	AA768437 ob22f07.s
	32	17.8	80.9	389	1	AA535707	AA535707 nf88d04.s
c	33	17.8	80.9	398	2	AW291203	AW291203 UI-H-BI2-
	34	17.8	80.9	404	1	AI242291	AI242291 qi16g02.x
	35	17.8	80.9	416	1	AI074659	AI074659 ox82g07.s
	36	17.8	80.9	423	7	CN545186	CN545186 EST 17130
c	37	17.8	80.9	428	1	AI041258	AI041258 ov66a02.x
	38	17.8	80.9	435	5	BM998692	BM998692 UI-H-DT1-
	39	17.8	80.9	444	1	AA417921	AA417921 zv94b03.s
c	40	17.8	80.9	447	1	AA834127	AA834127 of26g07.s
	41	17.8	80.9	447	2	AW073711	AW073711 xb01h10.x
c	42	17.8	80.9	448	1	AI418470	AI418470 tg48e05.x
	43	17.8	80.9	450	1	AA768438	AA768438 ob22f08.s
c	44	17.8	80.9	450	1	AA418172	AA418172 zv94f03.r
c	45	17.8	80.9	452	5	BQ942517	BQ942517 AGENCOURT
c	46	17.8	80.9	455	5	BP305540	BP305540 BP305540
c	47	17.8	80.9	457	6	BY560157	BY560157 BY560157
c	48	17.8	80.9	460	2	BE379976	BE379976 601159478
	49	17.8	80.9	463	1	AA159601	AA159601 zo80a04.s
c	50	17.8	80.9	463	2	AW275079	AW275079 xn82c01.x
	51	17.8	80.9	470	5	BQ012758	BQ012758 UI-1-BC1p
	52	17.8	80.9	484	1	AI088652	AI088652 qb14a07.x
	53	17.8	80.9	490	1	AA936255	AA936255 on75b04.s
c	54	17.8	80.9	494	2	AW471206	AW471206 xv13e09.x
c	55	17.8	80.9	506	5	BX279662	BX279662 BX279662
	56	17.8	80.9	516	5	BQ774467	BQ774467 UI-H-EZ1-
c	57	17.8	80.9	526	2	BE440143	BE440143 HTM1-954R
	58	17.8	80.9	530	1	AI400162	AI400162 tg67g09.x
	59	17.8	80.9	534	2	AW128873	AW128873 xe89c02.x
	60	17.8	80.9	541	1	AA864874	AA864874 oh03d08.s
c	61	17.8	80.9	546	2	BF890876	BF890876 PM2-MT010
c	62	17.8	80.9	547	4	BI962926	BI962926 ie60d12.y
	63	17.8	80.9	549	4	BI966885	BI966885 ie62f08.x
c	64	17.8	80.9	550	4	BI963087	BI963087 ie62f08.y
	65	17.8	80.9	551	1	AI088606	AI088606 qb14e01.x
c	66	17.8	80.9	555	1	AV735110	AV735110 AV735110
c	67	17.8	80.9	558	2	AW970416	AW970416 EST382497
c	68	17.8	80.9	563	6	CB267636	CB267636 1006542 H
c	69	17.8	80.9	568	1	AV715652	AV715652 AV715652
c	70	17.8	80.9	568	4	BF970233	BF970233 602273470
	71	17.8	80.9	568	4	BG723880	BG723880 602697376
c	72	17.8	80.9	569	1	AL701719	AL701719 DKFZp686H
	73	17.8	80.9	569	5	BX094624	BX094624 BX094624
	74	17.8	80.9	576	4	BM670715	BM670715 UI-E-DX1-
	75	17.8	80.9	577	1	AI309768	AI309768 qo75b03.x
c	76	17.8	80.9	582	5	BP379530	BP379530 BP379530
	77	17.8	80.9	584	4	BI966756	BI966756 ie60d12.x
c	78	17.8	80.9	585	6	CB160326	CB160326 K-EST0220
c	79	17.8	80.9	593	2	BE739363	BE739363 601556586
c	80	17.8	80.9	613	1	AV716017	AV716017 AV716017
c	81	17.8	80.9	618	7	CV027350	CV027350 5580 Full
c	82	17.8	80.9	620	4	BG500991	BG500991 602546583
	83	17.8	80.9	623	2	BE739605	BE739605 601556586
c	84	17.8	80.9	643	4	BI553318	BI553318 603193449
c	85	17.8	80.9	647	4	BG535175	BG535175 602562794
c	86	17.8	80.9	654	4	BG505886	BG505886 601859895
c	87	17.8	80.9	657	4	BM713190	BM713190 UI-E-EJ0-
c	88	17.8	80.9	657	8	B63640	B63640 RPCI11-2J13
c	89	17.8	80.9	663	2	BF213738	BF213738 601847628
c	90	17.8	80.9	663	6	CD639326	CD639326 AGENCOURT
	91	17.8	80.9	673	9	CE750532	CE750532 tigr-gss-
c	92	17.8	80.9	682	6	CD641555	CD641555 AGENCOURT
c	93	17.8	80.9	721	1	AV762336	AV762336 AV762336
c	94	17.8	80.9	738	2	BE603221	BE603221 HVSMEh010
c	95	17.8	80.9	742	1	AV758775	AV758775 AV758775
c	96	17.8	80.9	765	4	BI562028	BI562028 603255083
c	97	17.8	80.9	766	2	BE566608	BE566608 601340137

98	17.8	80.9	766	9	BX204503	BX204503 Danio rer
99	17.8	80.9	775	2	BF529524	BF529524 602043291
c 100	17.8	80.9	785	1	AV757859	AV757859 AV757859

ALIGNMENTS

RESULT 1
CC229626
LOCUS CC229626 1174 bp DNA linear GSS 12-MAY-2003
DEFINITION CH261-46H2_Sp6.1 CH261 Gallus gallus genomic clone CH261-46H2, genomic survey sequence.
ACCESSION CC229626
VERSION CC229626.1 GI:30556289
KEYWORDS GSS.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 1174)
AUTHORS Kremitzki,C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J., Warren,W., Graves,T., Mardis,E. and Wilson,R.
TITLE Gallus gallus BAC End Reads
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 182000 Std Error: 0.00
Seq primer: Sp6 ATTTAGGTGACACTATAG
Class: BAC ends
High quality sequence start: 329
High quality sequence stop: 482.
FEATURES Location/Qualifiers
source 1. .1174
/organism="Gallus gallus"
/mol_type="genomic DNA"
/strain="Red Jungle Fowl"
/db_xref="taxon:9031"
/clone="CH261-46H2"
/sex="female"
/cell_line="UCD001, inbred 256"
/clone_lib="CH261"
/note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI; CH261 Female Chicken library - for library and clone ordering information: http://www.chori.org/bacpac"
ORIGIN

Query Match 86.4%; Score 19; DB 8; Length 1174;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          4 CCTCTTCAGTCCAATCAAG 22
              ||||||||||||||||||
Db        399 CCTCTTCAGTCCAATCAAG 417
```

RESULT 2
BE541043
LOCUS BE541043 767 bp mRNA linear EST 09-AUG-2000
DEFINITION 601064391F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:3450647 5', mRNA sequence.
ACCESSION BE541043
VERSION BE541043.1 GI:9769787
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 767)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM8429 row: f column: 24
High quality sequence stop: 523.
FEATURES Location/Qualifiers
source 1. .767
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3450647"
/cell_line="MGC36"
/lab_host="DH10B"
/clone_lib="NIH_MGC_10"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.5 kb. Library prepared by Life Technologies."
ORIGIN

Query Match 85.5%; Score 18.8; DB 2; Length 767;
Best Local Similarity 90.9%; Pred. No. 4.5e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              || |||||||||| ||||||||
Db        107 AGTCCTCTTCAGTGCAATCAAG 128
```

RESULT 3
BF687801
LOCUS BF687801 872 bp mRNA linear EST 22-DEC-2000
DEFINITION 602066853F1 NIH_MGC_57 Homo sapiens cDNA clone IMAGE:4065901 5', mRNA sequence.
ACCESSION BF687801
VERSION BF687801.1 GI:11973209
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 872)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM902 row: j column: 14
High quality sequence stop: 623.
FEATURES Location/Qualifiers
source 1. .872
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4065901"
/tissue_type="glioblastoma"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_57"

/note="Organ: brain; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.55 kb (range 0.9-4.0 kb). 12/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

ORIGIN

```
  Query Match             85.5%;  Score 18.8;  DB 2;  Length 872;
  Best Local Similarity   90.9%;  Pred. No. 4.5e+02;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              || |||||||||| ||||||||
Db        319 AGTCCTCTTCAGTGCAATCAAG 340
```

RESULT 4
CL901625/c
LOCUS CL901625 935 bp DNA linear GSS 30-AUG-2004
DEFINITION CSHC1000 1639HC library Triticum aestivum genomic clone 1639HC06E05, genomic survey sequence.
ACCESSION CL901625
VERSION CL901625.1 GI:51663670
KEYWORDS GSS.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 935)
AUTHORS Lamoureux,D., Peterson,D.G., Li,W., Fellers,J.P. and Gill,B.S.
TITLE C0t-based cloning and sequencing (CBCS) efficiently removes sequence repeats and increases gene ratio in bread wheat
JOURNAL Unpublished (2004)
COMMENT Contact: Gill BS
Department of Plant Pathology
Kansas State University
4024 Throckmorton, Manhattan, KS 66506-5502, USA
Tel: 785 532 1391
Fax: 785 532 5692
Email: bsgill@ksu.edu
Seq primer: T7
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .935
/organism="Triticum aestivum"
/mol_type="genomic DNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="1639HC06E05"
/tissue_type="whole plant"
/dev_stage="young shoot"
/clone_lib="1639HC library"

ORIGIN

```
  Query Match             83.6%;  Score 18.4;  DB 9;  Length 935;
  Best Local Similarity   95.0%;  Pred. No. 7.1e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 CCCTCTTCAGTCCAATCAAG 22
              |||||||||||| |||||||
Db        216 CCCTCTTCAGTCAAATCAAG 197
```

RESULT 5
AA693379/c
LOCUS AA693379 169 bp mRNA linear EST 12-JAN-1999
DEFINITION ah21g06.s1 Soares_parathyroid_tumor_NbHPA Homo sapiens cDNA clone 1239514 3' similar to TR:Q13227 Q13227 GPS2. ;, mRNA sequence.
ACCESSION AA693379
VERSION AA693379.1 GI:2694317
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 169)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality
Possible reversed clone: similarity on wrong strand
Insert Length: 912 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 1.
FEATURES Location/Qualifiers
source 1. .169
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="1239514"
/tissue_type="parathyroid tumor"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares_parathyroid_tumor_NbHPA"
/note="Organ: parathyroid gland; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCACCAATTTTTTTTTTTTTTTTTT TTTTT-3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. RNA from sporadic parathyroid adenomas was kindly provided by Dr. Stephen Marx, National Institute of Diabetes and Digestive and Kidney Diseases, NIH."

ORIGIN

```
  Query Match             80.9%;  Score 17.8;  DB 1;  Length 169;
  Best Local Similarity   90.5%;  Pred. No. 1.1e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 GCCCTCTTCAGTCCAATCAAG 22
              ||| ||||||| |||||||||
Db        130 GCCATCTTCAGCCCAATCAAG 110
```

RESULT 6
BF833890/c
LOCUS BF833890 185 bp mRNA linear EST 13-JAN-2001
DEFINITION RC1-HT0881-041100-019-b04 HT0881 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF833890
VERSION BF833890.1 GI:12183723
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 185)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
            O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=RC1&t2=RC1-HT0881-
            041100-019-b04&t3=2000-11-04&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 185.
FEATURES             Location/Qualifiers
     source          1. .185
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="HT0881"
                     /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
ORIGIN

  Query Match             80.9%;  Score 17.8;  DB 2;  Length 185;
  Best Local Similarity   90.5%;  Pred. No. 1.1e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 GCCCTCTTCAGTCCAATCAAG 22
              ||| |||||||||||||| ||
Db         29 GCCATCTTCAGTCCAATCCAG 9


RESULT 7
AW800621
LOCUS       AW800621                193 bp    mRNA    linear   EST 16-MAY-2000
DEFINITION  MR1-UM0063-080300-001-c08 UM0063 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW800621
VERSION     AW800621.1  GI:7852491
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 193)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
            O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=MR1-UM0063-080
            300-001-c08&t3=2000-03-08&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 14
            High quality sequence stop: 193.
FEATURES             Location/Qualifiers
     source          1. .193
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="UM0063"
                     /note="Organ: uterus; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No.
                     196,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
ORIGIN

  Query Match             80.9%;  Score 17.8;  DB 2;  Length 193;
  Best Local Similarity   90.5%;  Pred. No. 1.1e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 GCCCTCTTCAGTCCAATCAAG 22
              ||| |||||||||||||| ||
Db         98 GCCATCTTCAGTCCAATCCAG 118


RESULT 8
BF111499
LOCUS       BF111499                205 bp    mRNA    linear   EST 20-OCT-2000
DEFINITION  7l28f04.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
            IMAGE:3522942 3' similar to TR:Q92478 Q92478 C-TYPE LECTIN. [1] ;,
            mRNA sequence.
ACCESSION   BF111499
VERSION     BF111499.1  GI:10941189
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 205)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Trace considered overall poor quality
            Seq primer: -40UP from Gibco
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .205
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
```

/clone="IMAGE:3522942"
/lab_host="DH10B"
/clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from five normalized libraries were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 5 libraries. The pools consisted of the following libraries and cloneIDs: Soares NbHSF pool 1: 309384-310919, 323208-325895 Soares Nb2HP pool 1: 145032-147335, 147720-148103, 148872-149255, 15002 - 150407, 151176-152327 Soares Nb2HF8-9W pool 1: 758280-760583, 772104-774407 Soares NbHPA pool 1: 304776-306311, 320136-322823, 326280-326663 Soares NbHOT pool 1: 723720-726407, 739080-740999 Subtraction by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 80.9%; Score 17.8; DB 2; Length 205;
Best Local Similarity 90.5%; Pred. No. 1.1e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GCCCTCTTCAGTCCAATCAAG 22
||| |||||||||||||| ||
Db 30 GCCATCTTCAGTCCAATCCAG 50

RESULT 9
AA884812/c
LOCUS AA884812 270 bp mRNA linear EST 24-OCT-2000
DEFINITION am28b05.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:1468113 3', mRNA sequence.
ACCESSION AA884812
VERSION AA884812.1 GI:2994793
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 270)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1662 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 153
POLYA=No.
FEATURES Location/Qualifiers
source 1. .270
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1468113"
/lab_host="DH10B"
/clone_lib="Soares_NFL_T_GBC_S1"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NbHL19W, testis NHT, and B-cell NCI_CGAP_GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo. "

ORIGIN

Query Match 80.9%; Score 17.8; DB 1; Length 270;
Best Local Similarity 90.5%; Pred. No. 1.2e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GCCCTCTTCAGTCCAATCAAG 22
|| |||||||||||||||||
Db 249 GCTGTCTTCAGTCCAATCAAG 229

RESULT 10
AW819544/c
LOCUS AW819544 270 bp mRNA linear EST 17-MAY-2000
DEFINITION RC5-ST0293-140200-013-E04 ST0293 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW819544
VERSION AW819544.1 GI:7912538
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 270)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=RC5-ST0293-140200-013-E04&t3=2000-02-14&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 270.
FEATURES Location/Qualifiers
source 1. .270
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="ST0293"
/note="Organ: stomach; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 80.9%; Score 17.8; DB 2; Length 270;
Best Local Similarity 90.5%; Pred. No. 1.2e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GCCCTCTTCAGTCCAATCAAG 22
||| |||||||||||||| ||
Db 138 GCCATCTTCAGTCCAATCCAG 118

```
RESULT 11
AA437133/c
LOCUS       AA437133                271 bp    mRNA    linear   EST 30-MAY-1997
DEFINITION  zv53e10.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:757386
            3', mRNA sequence.
ACCESSION   AA437133
VERSION     AA437133.1  GI:2142047
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 271)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
            Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -41m13 fwd. ET from Amersham
            High quality sequence stop: 166.
FEATURES             Location/Qualifiers
     source          1. .271
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:5978000"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:757386"
                     /sex="male"
                     /lab_host="DH10B"
                     /clone_lib="Soares_testis_NHT"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech
                     Laboratories, Inc., and primed with  a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
ORIGIN

  Query Match             80.9%;  Score 17.8;  DB 1;  Length 271;
  Best Local Similarity   90.5%;  Pred. No. 1.2e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 GCCCTCTTCAGTCCAATCAAG 22
              ||  ||||||||||||||||||
Db        251 GCTGTCTTCAGTCCAATCAAG 231


RESULT 12
AW467041
LOCUS       AW467041                271 bp    mRNA    linear   EST 24-FEB-2000
DEFINITION  ha08c03.x1 NCI_CGAP_Kid12 Homo sapiens cDNA clone IMAGE:2873092 3'
            similar to TR:Q92478 Q92478 C-TYPE LECTIN. ;, mRNA sequence.
ACCESSION   AW467041
VERSION     AW467041.1  GI:7037147
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 271)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Seq primer: -40UP from Gibco
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .271
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2873092"
                     /tissue_type="2 pooled tumors (clear cell type)"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Kid12"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Plasmid DNA from the normalized library NCI_CGAP_Kid5 was
                     prepared, and ss circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1323912-1325831, 1471368-1472903 and
                     1492104-1493255). Subtraction by Bento Soares and M.
                     Fatima Bonaldo.  "
ORIGIN

  Query Match             80.9%;  Score 17.8;  DB 2;  Length 271;
  Best Local Similarity   90.5%;  Pred. No. 1.2e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 GCCCTCTTCAGTCCAATCAAG 22
              ||| ||||||||||||||| ||
Db        153 GCCATCTTCAGTCCAATCCAG 173


RESULT 13
BB720494
LOCUS       BB720494                289 bp    mRNA    linear   EST 12-OCT-2001
DEFINITION  BB720494 RIKEN full-length enriched, adult male liver tumor Mus
            musculus cDNA clone C730036G02 3', mRNA sequence.
ACCESSION   BB720494
VERSION     BB720494.1  GI:16102067
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 289)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K.,
            Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
            Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
            Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
```

```
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
            and Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
            Sugahara,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .289
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="C730036G02"
                     /sex="male"
                     /tissue_type="liver tumor"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, adult male liver
                     tumor"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA
                     was cleaved with BamHI and XhoI. Vector: a modified
                     pBluescript KS(+) after bulk excision from Lambda FLC I.
                     Tissue was provided by William A. Held, Roswell Park
                     Cancer Institute, Department of Molecular and Cellular
                     Biology, Elm and Carlton Streets, Buffalo, NY 14263, whose
                     assistance we gratefully acknowledge."
ORIGIN

  Query Match             80.9%;  Score 17.8;  DB 2;  Length 289;
  Best Local Similarity   90.5%;  Pred. No. 1.2e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAA 21
              ||||||||||| |||| ||||
Db         10 AGCCCTCTTCATTCCACTCAA 30
```

```
RESULT 14
BM151302
LOCUS       BM151302                295 bp    mRNA    linear   EST 30-NOV-2001
DEFINITION  TCBAP1D6134 Pediatric pre-B cell acute lymphoblastic leukemia
            Baylor-HGSC project=TCBA Homo sapiens cDNA clone TCBAP6134, mRNA
            sequence.
ACCESSION   BM151302
VERSION     BM151302.1  GI:17174603
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 295)
  AUTHORS   Wei,Y., Tsang,Y.T.M., Mei,G., Ku,J.M., Ali-Osman,F.R. Jr.,
            Gunaratne,P.H., Muzny,D., Bouck,J., Gibbs,R.A. and Margolin,J.F.
  TITLE     Pediatric Leukemia cDNA Sequencing Project (2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Dr. Judith F. Margolin
            Texas Children's Cancer Center and Human Genome Sequencing Center
            at Baylor College of Medicine
            1102 Bates, MC3-3320 Houston, TX 77030, USA
            Tel: 832-824-4536
            Fax: 832-825-4038
            Email: clones@txccc.org
            Seq primer: M13 primer.
FEATURES             Location/Qualifiers
     source          1. .295
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="TCBAP6134"
                     /sex="male"
                     /tissue_type="leukopheresis"
                     /cell_type="pre-B cell"
                     /dev_stage="pediatric 2 years"
                     /lab_host="DH10B"
                     /clone_lib="Pediatric pre-B cell acute lymphoblastic
                     leukemia Baylor-HGSC project=TCBA"
                     /note="Vector: lambda pSB; Site_1: BamHI; Site_2: EcoRI;
                     First strand cDNA was primed with an anchored
                     XhoI-oligo(dT) primer [5'GGAGGACTCGAGCGGCCGCAGGAGGAG(T)VN
                     3'; V=A,C,G; N=A,C,G,T] and then dG tailed. Second strand
                     was primed with a BamH1-dC primer
                     [5'AGAGAGCTCGGATCCGCGGCCGCAATAATAATAAT(C) 3'].
                     Double-stranded cDNA was then digested with BamH1 and XhoI
                     and directionally cloned into the BamH1 and SalI sites of
                     lambda pSB vector. Library went through one round of
                     normalization. Library was constructed by Wei Yu at RIKEN
                     of Japan (Carninci P, Westover A, Nishiyama Y, Ohsumi T,
                     Itoh M, Nagaoka S, SasakiN, Okazaki Y, Muramatsu M,
                     Schneider C, Hayashizaki Y, High efficiency selection of
                     full-length cDNA by improved biotinylated cap trapper.,
                     DNA Res 4: 1, 61-6, Feb 28, 1997)"
ORIGIN

  Query Match             80.9%;  Score 17.8;  DB 4;  Length 295;
  Best Local Similarity   90.5%;  Pred. No. 1.2e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 GCCCTCTTCAGTCCAATCAAG 22
              ||| |||||||||||||| ||
Db        269 GCCATCTTCAGTCCAATCCAG 289


RESULT 15
AI203209/c
LOCUS       AI203209                309 bp    mRNA    linear   EST 03-FEB-1999
DEFINITION  qr23h08.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:1941759 3',
            mRNA sequence.
ACCESSION   AI203209
VERSION     AI203209.1  GI:3755815
KEYWORDS    EST.
```

```
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 309)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 751   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 295.
FEATURES             Location/Qualifiers
     source          1. .309
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1941759"
                     /tissue_type="pooled germ cell tumors"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_GC6"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA
                     from the normalized library NCI_CGAP_GC4 was prepared, and
                     ss circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (cloneIDs
                     1257096-1258631, 1469064-1470983, and 1475592-1476743).
                     Subtraction by Bento Soares and M. Fatima Bonaldo.  "
ORIGIN

  Query Match             80.9%;  Score 17.8;  DB 1;  Length 309;
  Best Local Similarity   90.5%;  Pred. No. 1.2e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 GCCCTCTTCAGTCCAATCAAG 22
              ||  ||||||||||||||||||
Db        244 GCTGTCTTCAGTCCAATCAAG 224


RESULT 16
T46996
LOCUS       T46996                  310 bp    mRNA    linear   EST 01-FEB-1995
DEFINITION  yb12b06.s1 Stratagene placenta (#937225) Homo sapiens cDNA clone
            IMAGE:70931 3' similar to similar to SP:CD69_HUMAN Q07108 EARLY
            ACTIVATION ANTIGEN CD69, mRNA sequence.
ACCESSION   T46996
VERSION     T46996.1  GI:648979
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 310)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
   PUBMED   8889549
COMMENT     Other_ESTs: yb12b06.r1
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 477
            High qality sequence stops: 272 Source: IMAGE Consortium, LLNL This
            clone is available royalty-free through LLNL ; contact the IMAGE
            Consortium (info@image.llnl.gov) for further information. Possible
            reversed clone: polyT not found
            Insert Length: 477   Std Error: 0.00
            Seq primer: -21m13
            High quality sequence stop: 272.
FEATURES             Location/Qualifiers
     source          1. .310
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:491828"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:70931"
                     /sex="male"
                     /lab_host="SOLR cells (kanamycin resistant)"
                     /clone_lib="Stratagene placenta (#937225)"
                     /note="Organ: placenta; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. Caucasian. Average insert size: 1.2 kb; Uni-ZAP
                     XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3'
                     adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
ORIGIN

  Query Match             80.9%;  Score 17.8;  DB 7;  Length 310;
  Best Local Similarity   90.5%;  Pred. No. 1.2e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 GCCCTCTTCAGTCCAATCAAG 22
              ||| ||||||||||||||  ||
Db        175 GCCATCTTCAGTCCAATCCAG 195


RESULT 17
AA969466
LOCUS       AA969466                328 bp    mRNA    linear   EST 07-JUL-1998
DEFINITION  oo81d01.s1 NCI_CGAP_Kid5 Homo sapiens cDNA clone IMAGE:1572577 3'
            similar to TR:Q92478 Q92478 C-TYPE LECTIN. ;, mRNA sequence.
ACCESSION   AA969466
VERSION     AA969466.1  GI:3144646
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 328)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
```

```
            Trace considered overall poor quality
            Insert Length: 1094   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 1.
FEATURES              Location/Qualifiers
     source           1. .328
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1572577"
                      /tissue_type="2 pooled tumors (clear cell type)"
                      /lab_host="DH10B"
                      /clone_lib="NCI_CGAP_Kid5"
                      /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                      a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer [5'
                      AACTGGAAGAATTCGCGGCCGCAATATTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization. Library
                      constructed by Bento Soares and M. Fatima Bonaldo.  "
ORIGIN

  Query Match             80.9%;  Score 17.8;  DB 1;  Length 328;
  Best Local Similarity   90.5%;  Pred. No. 1.2e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 GCCCTCTTCAGTCCAATCAAG 22
              ||| |||||||||||||| ||
Db        153 GCCATCTTCAGTCCAATCCAG 173


RESULT 18
AW615159/c
LOCUS       AW615159                328 bp    mRNA    linear   EST 23-MAR-2000
DEFINITION  hg73h02.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2951283 3',
            mRNA sequence.
ACCESSION   AW615159
VERSION     AW615159.1  GI:7320345
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 328)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            image.llnl.gov/image/html/iresources.shtml
            Seq primer: -40UP from Gibco.
FEATURES              Location/Qualifiers
     source           1. .328
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:2951283"
                      /tissue_type="pooled germ cell tumors"
                      /lab_host="DH10B"
                      /clone_lib="NCI_CGAP_GC6"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA
                      from the normalized library NCI_CGAP_GC4 was prepared, and
                      ss circles were made in vitro. Following HAP purification,
                      this DNA was used as tracer in a subtractive hybridization
                      reaction. The driver was PCR-amplified cDNAs from a pool
                      of 5,000 clones made from the same library (cloneIDs
                      1257096-1258631, 1469064-1470983, and 1475592-1476743).
                      Subtraction by Bento Soares and M. Fatima Bonaldo.  "
ORIGIN

  Query Match             80.9%;  Score 17.8;  DB 2;  Length 328;
  Best Local Similarity   90.5%;  Pred. No. 1.2e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 GCCCTCTTCAGTCCAATCAAG 22
              ||  ||||||||||||||||||
Db        230 GCTGTCTTCAGTCCAATCAAG 210


RESULT 19
AI307235
LOCUS       AI307235                339 bp    mRNA    linear   EST 08-APR-1999
DEFINITION  tb18c07.x1 NCI_CGAP_Kid12 Homo sapiens cDNA clone IMAGE:2054700 3'
            similar to TR:Q92478 Q92478 C-TYPE LECTIN. ;, mRNA sequence.
ACCESSION   AI307235
VERSION     AI307235.1  GI:4001991
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 339)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 603   Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES              Location/Qualifiers
     source           1. .339
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:2054700"
                      /tissue_type="2 pooled tumors (clear cell type)"
                      /lab_host="DH10B"
                      /clone_lib="NCI_CGAP_Kid12"
                      /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                      a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                      Plasmid DNA from the normalized library NCI_CGAP_Kid5 was
                      prepared, and ss circles were made in vitro. Following HAP
                      purification, this DNA was used as tracer in a subtractive
                      hybridization reaction. The driver was PCR-amplified cDNAs
                      from a pool of 5,000 clones made from the same library
                      (cloneIDs 1323912-1325831, 1471368-1472903 and
                      1492104-1493255). Subtraction by Bento Soares and M.
                      Fatima Bonaldo.  "
ORIGIN

  Query Match             80.9%;  Score 17.8;  DB 1;  Length 339;
  Best Local Similarity   90.5%;  Pred. No. 1.2e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 GCCCTCTTCAGTCCAATCAAG 22
```

```
                ||| |||||||||||||| ||
Db        254 GCCATCTTCAGTCCAATCCAG 274


RESULT 20
AI631503/c
LOCUS       AI631503                339 bp    mRNA    linear   EST 16-DEC-1999
DEFINITION  wa89g09.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2303392 3',
            mRNA sequence.
ACCESSION   AI631503
VERSION     AI631503.1  GI:4682833
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 339)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 424   Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .339
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2303392"
                     /tissue_type="pooled germ cell tumors"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_GC6"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA
                     from the normalized library NCI_CGAP_GC4 was prepared, and
                     ss circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (cloneIDs
                     1257096-1258631, 1469064-1470983, and 1475592-1476743).
                     Subtraction by Bento Soares and M. Fatima Bonaldo.  "
ORIGIN

  Query Match             80.9%;  Score 17.8;  DB 1;  Length 339;
  Best Local Similarity   90.5%;  Pred. No. 1.2e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 GCCCTCTTCAGTCCAATCAAG 22
              ||  ||||||||||||||||||
Db        244 GCTGTCTTCAGTCCAATCAAG 224


RESULT 21
AI633750/c
LOCUS       AI633750                339 bp    mRNA    linear   EST 14-DEC-1999
DEFINITION  tt28b04.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2242063 3',
            mRNA sequence.
ACCESSION   AI633750
VERSION     AI633750.1  GI:4685080
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 339)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 422   Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .339
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2242063"
                     /tissue_type="pooled germ cell tumors"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_GC6"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA
                     from the normalized library NCI_CGAP_GC4 was prepared, and
                     ss circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (cloneIDs
                     1257096-1258631, 1469064-1470983, and 1475592-1476743).
                     Subtraction by Bento Soares and M. Fatima Bonaldo.  "
ORIGIN

  Query Match             80.9%;  Score 17.8;  DB 1;  Length 339;
  Best Local Similarity   90.5%;  Pred. No. 1.2e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 GCCCTCTTCAGTCCAATCAAG 22
              ||  ||||||||||||||||||
Db        244 GCTGTCTTCAGTCCAATCAAG 224


RESULT 22
AI954798/c
LOCUS       AI954798                339 bp    mRNA    linear   EST 06-SEP-1999
DEFINITION  wq33d07.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2473069 3',
            mRNA sequence.
ACCESSION   AI954798
VERSION     AI954798.1  GI:5747108
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 339)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
```

DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco.

FEATURES Location/Qualifiers
source 1. .339
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2473069"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_GC6"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo. "

ORIGIN

Query Match 80.9%; Score 17.8; DB 1; Length 339;
Best Local Similarity 90.5%; Pred. No. 1.2e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GCCCTCTTCAGTCCAATCAAG 22
|| ||||||||||||||||||
Db 244 GCTGTCTTCAGTCCAATCAAG 224

RESULT 23
AI954834/c
LOCUS AI954834 339 bp mRNA linear EST 08-MAR-2000
DEFINITION wq33h04.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2473111 3', mRNA sequence.
ACCESSION AI954834
VERSION AI954834.1 GI:5747144
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 339)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 390 Std Error: 0.00
Seq primer: -40UP from Gibco.

FEATURES Location/Qualifiers
source 1. .339
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2473111"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_GC6"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo. "

ORIGIN

Query Match 80.9%; Score 17.8; DB 1; Length 339;
Best Local Similarity 90.5%; Pred. No. 1.2e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GCCCTCTTCAGTCCAATCAAG 22
|| ||||||||||||||||||
Db 244 GCTGTCTTCAGTCCAATCAAG 224

RESULT 24
AI962281/c
LOCUS AI962281 339 bp mRNA linear EST 08-MAR-2000
DEFINITION wq46e04.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2474334 3', mRNA sequence.
ACCESSION AI962281
VERSION AI962281.1 GI:5754994
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 339)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 382 Std Error: 0.00
Seq primer: -40UP from Gibco.

FEATURES Location/Qualifiers
source 1. .339
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2474334"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_GC6"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo. "

ORIGIN

Query Match 80.9%; Score 17.8; DB 1; Length 339;
Best Local Similarity 90.5%; Pred. No. 1.2e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          2 GCCCTCTTCAGTCCAATCAAG 22
              ||  ||||||||||||||||||
Db        244 GCTGTCTTCAGTCCAATCAAG 224


RESULT 25
T46995/c
LOCUS       T46995                  339 bp    mRNA    linear   EST 01-FEB-1995
DEFINITION  yb12b06.r1 Stratagene placenta (#937225) Homo sapiens cDNA clone
            IMAGE:70931 5' similar to similar to SP:CD69_HUMAN Q07108 EARLY
            ACTIVATION ANTIGEN CD69, mRNA sequence.
ACCESSION   T46995
VERSION     T46995.1  GI:648978
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 339)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
   PUBMED   8889549
COMMENT     Other_ESTs: yb12b06.s1
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 477
            High qality sequence stops: 243 Source: IMAGE Consortium, LLNL This
            clone is available royalty-free through LLNL ; contact the IMAGE
            Consortium (info@image.llnl.gov) for further information.
            Insert Length: 477   Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 243.
FEATURES             Location/Qualifiers
     source          1. .339
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:491828"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:70931"
                     /sex="male"
                     /lab_host="SOLR cells (kanamycin resistant)"
                     /clone_lib="Stratagene placenta (#937225)"
                     /note="Organ: placenta; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. Caucasian. Average insert size: 1.2 kb; Uni-ZAP
                     XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3'
                     adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
ORIGIN

  Query Match             80.9%;  Score 17.8;  DB 7;  Length 339;
  Best Local Similarity   90.5%;  Pred. No. 1.2e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 GCCCTCTTCAGTCCAATCAAG 22
              ||| ||||||||||||| ||
Db        118 GCCATCTTCAGTCCAATCCAG 98


Search completed: September  6, 2005, 21:55:54
Job time : 1578.31 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 6, 2005, 19:12:21 ; Search time 62.9062 Seconds
(without alignments)
572.250 Million cell updates/sec

Title: US-10-729-421-35
Perfect score: 22
Sequence: 1 agccctcttcagtccaatcaag 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	17.8	80.9	247	4	US-09-513-999C-13311	Sequence 13311, A
	2	17.8	80.9	380	4	US-09-621-976-15744	Sequence 15744, A
c	3	17.8	80.9	759	4	US-09-944-807-5	Sequence 5, Appli
	4	17.8	80.9	1664976	4	US-08-916-421B-1	Sequence 1, Appli
	5	17.8	80.9	1664976	4	US-09-692-570-1	Sequence 1, Appli
c	6	17.2	78.2	1855	3	US-08-961-083-71	Sequence 71, Appl
c	7	17.2	78.2	1855	4	US-09-536-784-71	Sequence 71, Appl
c	8	17.2	78.2	3615	4	US-09-583-110-1422	Sequence 1422, Ap
c	9	17.2	78.2	3789	4	US-09-107-433-2464	Sequence 2464, Ap
	10	17.2	78.2	4191	4	US-10-144-198-13	Sequence 13, Appl
c	11	17.2	78.2	16535	3	US-08-961-527-74	Sequence 74, Appl
	12	17.2	78.2	55328	4	US-09-949-016-12747	Sequence 12747, A
	13	17.2	78.2	55330	4	US-09-949-016-17146	Sequence 17146, A
c	14	16.4	74.5	963	3	US-09-092-437-1	Sequence 1, Appli
c	15	16.4	74.5	963	4	US-09-583-110-1197	Sequence 1197, Ap
c	16	16.4	74.5	963	4	US-09-107-433-533	Sequence 533, App
c	17	16.4	74.5	3694	4	US-09-324-258-6	Sequence 6, Appli
	18	16.4	74.5	3902	3	US-08-961-527-212	Sequence 212, App
	19	16.4	74.5	232547	4	US-09-949-016-16603	Sequence 16603, A
c	20	16.2	73.6	601	4	US-09-949-016-145791	Sequence 145791,
c	21	16.2	73.6	601	4	US-09-949-016-146059	Sequence 146059,
c	22	16.2	73.6	601	4	US-09-949-016-146327	Sequence 146327,
c	23	16.2	73.6	601	4	US-09-949-016-195128	Sequence 195128,
	24	16.2	73.6	627	4	US-09-533-029-9	Sequence 9, Appli
	25	16.2	73.6	850	4	US-09-640-211A-212	Sequence 212, App
c	26	16.2	73.6	2634	4	US-09-634-238-408	Sequence 408, App
c	27	16.2	73.6	34049	4	US-09-949-016-17276	Sequence 17276, A
c	28	16.2	73.6	38009	4	US-09-949-016-13617	Sequence 13617, A
	29	16.2	73.6	39443	4	US-09-949-016-14326	Sequence 14326, A
	30	16.2	73.6	39443	4	US-09-949-016-14327	Sequence 14327, A
	31	16.2	73.6	143550	4	US-09-949-016-14143	Sequence 14143, A
	32	16.2	73.6	175236	4	US-09-949-016-14353	Sequence 14353, A
	33	16.2	73.6	205044	4	US-09-949-016-15851	Sequence 15851, A
	34	16.2	73.6	205044	4	US-09-949-016-15852	Sequence 15852, A
	35	16.2	73.6	205044	4	US-09-949-016-15853	Sequence 15853, A
	36	16.2	73.6	223471	4	US-09-949-016-12387	Sequence 12387, A
	37	16.2	73.6	223471	4	US-09-949-016-12724	Sequence 12724, A
	38	16.2	73.6	223471	4	US-09-949-016-12725	Sequence 12725, A
	39	16.2	73.6	246230	4	US-09-949-016-17019	Sequence 17019, A
	40	16.2	73.6	246230	4	US-09-949-016-17020	Sequence 17020, A
	41	16.2	73.6	246230	4	US-09-949-016-17021	Sequence 17021, A
	42	16.2	73.6	246230	4	US-09-949-016-17022	Sequence 17022, A
	43	16	72.7	5203	4	US-09-949-016-13950	Sequence 13950, A
	44	16	72.7	25992	4	US-09-949-016-17308	Sequence 17308, A
	45	16	72.7	26000	4	US-09-843-376-10	Sequence 10, Appl
	46	16	72.7	301828	4	US-09-949-016-13969	Sequence 13969, A
c	47	15.8	71.8	601	4	US-09-949-016-66443	Sequence 66443, A
	48	15.8	71.8	601	4	US-09-949-016-188498	Sequence 188498,
	49	15.8	71.8	601	4	US-09-949-016-188499	Sequence 188499,
	50	15.8	71.8	601	4	US-09-949-016-188500	Sequence 188500,
	51	15.8	71.8	601	4	US-09-949-016-188501	Sequence 188501,
c	52	15.8	71.8	1206	4	US-09-489-039A-4304	Sequence 4304, Ap
c	53	15.8	71.8	1665	1	US-08-440-377A-1	Sequence 1, Appli
c	54	15.8	71.8	1665	1	US-08-687-852-1	Sequence 1, Appli
c	55	15.8	71.8	1668	4	US-09-489-039A-2731	Sequence 2731, Ap
	56	15.8	71.8	18319	4	US-09-949-016-17446	Sequence 17446, A
	57	15.8	71.8	63563	4	US-09-596-002-33	Sequence 33, Appl
c	58	15.8	71.8	110243	4	US-09-949-016-13698	Sequence 13698, A
	59	15.8	71.8	148567	4	US-09-801-876B-3	Sequence 3, Appli
	60	15.8	71.8	148567	4	US-10-254-869-3	Sequence 3, Appli
	61	15.8	71.8	148567	4	US-10-667-442-3	Sequence 3, Appli
	62	15.8	71.8	158735	4	US-09-949-016-11989	Sequence 11989, A
	63	15.8	71.8	158735	4	US-09-949-016-17130	Sequence 17130, A
c	64	15.8	71.8	198632	4	US-09-949-016-12781	Sequence 12781, A
c	65	15.8	71.8	198637	4	US-09-949-016-17393	Sequence 17393, A
c	66	15.6	70.9	382	4	US-09-513-999C-30372	Sequence 30372, A
	67	15.6	70.9	601	4	US-09-949-016-149577	Sequence 149577,
c	68	15.6	70.9	812	4	US-09-597-771-13	Sequence 13, Appl
	69	15.6	70.9	921	4	US-09-614-221A-328	Sequence 328, App
	70	15.6	70.9	1842	3	US-09-832-498-1	Sequence 1, Appli
	71	15.6	70.9	1842	3	US-09-832-614A-1	Sequence 1, Appli
	72	15.6	70.9	2055	4	US-09-946-678-1	Sequence 1, Appli
	73	15.6	70.9	3151	4	US-09-917-254-43	Sequence 43, Appl
c	74	15.6	70.9	4944	1	US-08-623-679-6	Sequence 6, Appli
c	75	15.6	70.9	4944	3	US-08-933-774-6	Sequence 6, Appli
c	76	15.6	70.9	4944	3	US-09-181-030-6	Sequence 6, Appli
c	77	15.6	70.9	4944	3	US-09-534-242-6	Sequence 6, Appli
c	78	15.6	70.9	4944	3	US-09-454-854-6	Sequence 6, Appli
c	79	15.6	70.9	4944	3	US-09-164-671-6	Sequence 6, Appli
c	80	15.6	70.9	4944	4	US-09-182-113-6	Sequence 6, Appli
c	81	15.6	70.9	4944	4	US-08-862-442-6	Sequence 6, Appli
c	82	15.6	70.9	5055	1	US-08-623-679-8	Sequence 8, Appli
c	83	15.6	70.9	5055	3	US-08-933-774-8	Sequence 8, Appli
c	84	15.6	70.9	5055	3	US-09-181-030-8	Sequence 8, Appli
c	85	15.6	70.9	5055	3	US-09-534-242-8	Sequence 8, Appli
c	86	15.6	70.9	5055	3	US-09-454-854-8	Sequence 8, Appli
c	87	15.6	70.9	5055	3	US-09-164-671-8	Sequence 8, Appli
c	88	15.6	70.9	5055	4	US-09-182-113-8	Sequence 8, Appli
c	89	15.6	70.9	5055	4	US-08-862-442-8	Sequence 8, Appli
c	90	15.6	70.9	5262	4	US-09-248-796A-5051	Sequence 5051, Ap
c	91	15.6	70.9	5725	4	US-09-949-016-1964	Sequence 1964, Ap
c	92	15.6	70.9	6274	4	US-09-949-016-335	Sequence 335, App
c	93	15.6	70.9	12055	4	US-09-949-016-13233	Sequence 13233, A
c	94	15.6	70.9	39243	4	US-09-949-016-12316	Sequence 12316, A
c	95	15.6	70.9	39243	4	US-09-949-016-15443	Sequence 15443, A
c	96	15.6	70.9	49603	4	US-09-949-016-13706	Sequence 13706, A
c	97	15.6	70.9	53722	4	US-09-949-016-12077	Sequence 12077, A
c	98	15.6	70.9	156942	4	US-09-949-016-12227	Sequence 12227, A
c	99	15.6	70.9	156950	4	US-09-949-016-15946	Sequence 15946, A
	100	15.6	70.9	237863	4	US-09-949-016-13404	Sequence 13404, A

ALIGNMENTS

```
RESULT 1
US-09-513-999C-13311
; Sequence 13311, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Duclert, A.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
;  FILE REFERENCE: 59.US2.REG
;  CURRENT APPLICATION NUMBER: US/09/513,999C
;  CURRENT FILING DATE:  2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/122,487
;  PRIOR FILING DATE: 1999-02-26
;  NUMBER OF SEQ ID NOS: 36681
;  SOFTWARE: Patent.pm
; SEQ ID NO 13311
;   LENGTH: 247
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-513-999C-13311

  Query Match             80.9%;  Score 17.8;  DB 4;  Length 247;
  Best Local Similarity   90.5%;  Pred. No. 26;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 GCCCTCTTCAGTCCAATCAAG 22
              ||  ||||||||||||||||||
Db        164 GCTGTCTTCAGTCCAATCAAG 184


RESULT 2
US-09-621-976-15744
; Sequence 15744, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 15744
;   LENGTH: 380
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-621-976-15744

  Query Match             80.9%;  Score 17.8;  DB 4;  Length 380;
  Best Local Similarity   90.5%;  Pred. No. 29;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 GCCCTCTTCAGTCCAATCAAG 22
              ||  ||||||||||||||||||
Db        325 GCTGTCTTCAGTCCAATCAAG 345


RESULT 3
US-09-944-807-5/c
; Sequence 5, Application US/09944807
; Patent No. 6773895
; GENERAL INFORMATION:
;  APPLICANT: Boehringer Ingelheim Pharma KG
;  TITLE OF INVENTION: Method for identifying substances which positively
;  TITLE OF INVENTION:  influence inflammatory conditions of chronic
;  TITLE OF INVENTION:  inflammatory airway diseases
;  FILE REFERENCE: 082_00n
;  CURRENT APPLICATION NUMBER: US/09/944,807
;  CURRENT FILING DATE:  2001-08-31
;  PRIOR APPLICATION NUMBER: UK 0021484.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 759
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-944-807-5

  Query Match             80.9%;  Score 17.8;  DB 4;  Length 759;
  Best Local Similarity   90.5%;  Pred. No. 33;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 GCCCTCTTCAGTCCAATCAAG 22
              ||| ||||||||||||||| ||
Db        425 GCCATCTTCAGTCCAATCCAG 405


RESULT 4
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
;  APPLICANT: Bult et al.
;  TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococc
; Patent No. 6503729
;  TITLE OF INVENTION:  jannaschii
;  FILE REFERENCE: PB275
;  CURRENT APPLICATION NUMBER: US/08/916,421B
;  CURRENT FILING DATE:  1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/024,428
;  PRIOR FILING DATE: 1996-08-22
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1664976
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84808)..(84808)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84812)..(84812)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98120)..(98120)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98159)..(98159)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98239)..(98239)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98266)..(98266)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
```

; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1664854)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

Query Match 80.9%; Score 17.8; DB 4; Length 1664976;
Best Local Similarity 90.5%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AGCCCTCTTCAGTCCAATCAA 21
|||| |||||||||| |||||
Db 1438420 AGCCATCTTCAGTCCTATCAA 1438440

RESULT 5
US-09-692-570-1
; Sequence 1, Application US/09692570
; Patent No. 6797466
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococc

```
; Patent No. 6797466
;  TITLE OF INVENTION:  jannaschii
;  FILE REFERENCE: PB275C1
;  CURRENT APPLICATION NUMBER: US/09/692,570
;  CURRENT FILING DATE:  2003-01-14
;  PRIOR APPLICATION NUMBER: US 60/024,428
;  PRIOR FILING DATE: 1996-08-22
;  PRIOR APPLICATION NUMBER: US 08/916,421
;  PRIOR FILING DATE: 1997-08-22
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1664976
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84808)..(84808)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84812)..(84812)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98120)..(98120)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98159)..(98159)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98239)..(98239)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98266)..(98266)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98343)..(98343)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (103998)..(103998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (148948)..(148948)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (163385)..(163385)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (191989)..(191989)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (191995)..(191995)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (231980)..(231980)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (234187)..(234187)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (234220)..(234220)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (234814)..(234814)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (309398)..(309398)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (309418)..(309418)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (312837)..(312837)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (312993)..(312993)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (319226)..(319226)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (559167)..(559167)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (559241)..(559241)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (600992)..(600992)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (622708)..(622708)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (657081)..(657081)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (657203)..(657203)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (674435)..(674435)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (682442)..(682442)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (713652)..(713652)
```

```
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (741684)..(741684)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (779455)..(779455)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (779676)..(779676)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (855539)..(855539)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (871619)..(871619)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1084830)..(1084830)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1096846)..(1096846)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1119881)..(1119881)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1130881)..(1130881)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1310988)..(1310988)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1313224)..(1313224)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1349473)..(1349473)
;   OTHER INFORMATION: n equals a, t, c, or g

  Query Match             80.9%;  Score 17.8;  DB 4;  Length 1664976;
  Best Local Similarity   90.5%;  Pred. No. 1.4e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAA 21
              |||| |||||||||| |||||
Db    1438420 AGCCATCTTCAGTCCTATCAA 1438440


RESULT 6
US-08-961-083-71/c
; Sequence 71, Application US/08961083
; Patent No. 6159469
;  GENERAL INFORMATION:
;    APPLICANT:  Choi et. al.
;    TITLE OF INVENTION:  Streptococcus pneumoniae Antigens and Vaccines
;    NUMBER OF SEQUENCES:  452
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/961,083
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brookes, A. Anders
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB340P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  71:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1855 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-961-083-71

  Query Match             78.2%;  Score 17.2;  DB 3;  Length 1855;
  Best Local Similarity   86.4%;  Pred. No. 78;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              |||| | |||||||||| |||||
Db       1502 AGCCTTTTTCAGTCCATTCAAG 1481


RESULT 7
US-09-536-784-71/c
; Sequence 71, Application US/09536784
; Patent No. 6573082
;   GENERAL INFORMATION:
;       APPLICANT: Choi et. al.
;       TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
;       NUMBER OF SEQUENCES: 452
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Human Genome Sciences, Inc.
;            STREET: 9410 Key West Avenue
;            CITY: Rockville
;            STATE: Maryland
;            COUNTRY: USA
;            ZIP: 20850
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;            COMPUTER: HP Vectra 486/33
;            OPERATING SYSTEM: MSDOS version 6.2
;            SOFTWARE: ASCII Text
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/536,784
;            FILING DATE: 30-Oct-1997
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/961,083
;            FILING DATE: OCT-30-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Michelle S. Marks
;            REGISTRATION NUMBER: 41,971
;            REFERENCE/DOCKET NUMBER: PB340P3
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (301) 309-8504
;            TELEFAX: (301) 309-8512
;   INFORMATION FOR SEQ ID NO: 71:
;       SEQUENCE CHARACTERISTICS:
```

```
;             LENGTH: 1855 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 71:
US-09-536-784-71

  Query Match             78.2%;  Score 17.2;  DB 4;  Length 1855;
  Best Local Similarity   86.4%;  Pred. No. 78;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              |||| | ||||||||| |||||
Db       1502 AGCCTTTTTCAGTCCATTCAAG 1481


RESULT 8
US-09-583-110-1422/c
; Sequence 1422, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
;  APPLICANT:  Lynn Doucette-Stamm et al.
;  TITLE OF INVENTION:  Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
;  TITLE OF INVENTION:  Pneumoniae for Diagnostics and Therapeutics
;  FILE REFERENCE:  PATH00-07A
;  CURRENT APPLICATION NUMBER: US/09/583,110
;  CURRENT FILING DATE:  2000-05-26
;  PRIOR APPLICATION NUMBER:  US 09/107,433
;  PRIOR FILING DATE:  1998-06-30
;  PRIOR APPLICATION NUMBER:  US 60/085,131
;  PRIOR FILING DATE:  1998-05-12
;  PRIOR APPLICATION NUMBER:  US 60/051,553
;  PRIOR FILING DATE:  1997-07-02
;  NUMBER OF SEQ ID NOS:  5322
; SEQ ID NO 1422
;   LENGTH: 3615
;   TYPE: DNA
;   ORGANISM: Streptococcus pneumoniae
US-09-583-110-1422

  Query Match             78.2%;  Score 17.2;  DB 4;  Length 3615;
  Best Local Similarity   86.4%;  Pred. No. 89;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              |||| | ||||||||| |||||
Db       3181 AGCCTTTTTCAGTCCATTCAAG 3160


RESULT 9
US-09-107-433-2464/c
; Sequence 2464, Application US/09107433
; Patent No. 6800744
;   GENERAL INFORMATION:
;        APPLICANT: Lynn A Doucette-Stamm  and David Bush
;        TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
;                            SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE   FOR DIAGNO
;                            THERAPEUTICS
;        NUMBER OF SEQUENCES: 5206
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;             STREET: 100 Beaver Street
;             CITY: Waltham
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02354
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: CD/ROM ISO9660
;             COMPUTER: <Unknown>
;             OPERATING SYSTEM: <Unknown>
;             SOFTWARE: <Unknown>
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/107,433
;             FILING DATE: 30-Jun-1998
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/ 085131
;             FILING DATE: May 12, 1998
;             APPLICATION NUMBER: 60/051553
;             FILING DATE: July 2, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ariniello, Pamela Deneke
;             REGISTRATION NUMBER: 40,489
;             REFERENCE/DOCKET NUMBER: GTC-011
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (781)893-5007
;             TELEFAX: (781)893-8277
;   INFORMATION FOR SEQ ID NO: 2464:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 3789 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: circular
;        MOLECULE TYPE: DNA (genomic)
;        HYPOTHETICAL: NO
;        ANTI-SENSE: NO
;        ORIGINAL SOURCE:
;             ORGANISM: Streptococcus pneumoniae
;        FEATURE:
;             NAME/KEY:  misc_feature
;             LOCATION: (B) LOCATION 1...3789
;        SEQUENCE DESCRIPTION: SEQ ID NO: 2464:
US-09-107-433-2464

  Query Match             78.2%;  Score 17.2;  DB 4;  Length 3789;
  Best Local Similarity   86.4%;  Pred. No. 90;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              |||| | ||||||||| |||||
Db       3355 AGCCTTTTTCAGTCCATTCAAG 3334


RESULT 10
US-10-144-198-13
; Sequence 13, Application US/10144198
; Patent No. 6833247
; GENERAL INFORMATION:
;  APPLICANT: OriGene Technologies Inc
;  TITLE OF INVENTION: Regulated Prostate Cance Genes
;  FILE REFERENCE: 9U 105 R1
;  CURRENT APPLICATION NUMBER: US/10/144,198
;  CURRENT FILING DATE:  2002-05-14
;  NUMBER OF SEQ ID NOS: 44
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
;   LENGTH: 4191
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (121)..(3246)
US-10-144-198-13

  Query Match             78.2%;  Score 17.2;  DB 4;  Length 4191;
  Best Local Similarity   86.4%;  Pred. No. 92;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              | |||||||||| || ||||||
Db        831 ACCCCTCTTCAGCCCCATCAAG 852


RESULT 11
US-08-961-527-74/c
; Sequence 74, Application US/08961527
; Patent No. 6420135
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Charles Kunsch
;    TITLE OF INVENTION:  Streptococcus pneumoniae Polynucleotides and Sequences
;    NUMBER OF SEQUENCES:  391
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/961,527
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brookes, A. Anders
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB340P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  74:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  16535 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-961-527-74

  Query Match             78.2%;  Score 17.2;  DB 3;  Length 16535;
  Best Local Similarity   86.4%;  Pred. No. 1.2e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              |||| | ||||||||| |||||
Db       3310 AGCCTTTTTCAGTCCATTCAAG 3289


RESULT 12
US-09-949-016-12747
; Sequence 12747, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12747
;   LENGTH: 55328
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-12747

  Query Match             78.2%;  Score 17.2;  DB 4;  Length 55328;
  Best Local Similarity   86.4%;  Pred. No. 1.5e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              |||||||||| |||| ||| |
Db      17089 AGCCCTCTTCACTCCAGTCATG 17110


RESULT 13
US-09-949-016-17146
; Sequence 17146, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17146
;   LENGTH: 55330
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-17146

  Query Match             78.2%;  Score 17.2;  DB 4;  Length 55330;
  Best Local Similarity   86.4%;  Pred. No. 1.5e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              |||||||||| |||| ||| |
Db      17089 AGCCCTCTTCACTCCAGTCATG 17110


RESULT 14
US-09-092-437-1/c
; Sequence 1, Application US/09092437
; Patent No. 6190881
;  GENERAL INFORMATION:
;    APPLICANT:  Wilding, Edwina Imogen
;    APPLICANT:  Black, Michael T.
;    APPLICANT:  Shilling, Lisa K.
;    APPLICANT:  Kosmatka, Anna L.
;    APPLICANT:  Jaworski, Deborah D.
;    APPLICANT:  Wang, Min
;    TITLE OF INVENTION:  nrdF
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dechert, Price & Rhoads
;      STREET:  4000 Bell Atlantic Tower, 1717 Arch Stre
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103-2793
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows 95
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/092,437
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
```

```
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Falk, Stephen T
;       REGISTRATION NUMBER:  36,795
;       REFERENCE/DOCKET NUMBER:  GM10155
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  215-994-2488
;       TELEFAX:  215-994-2222
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  963 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
US-09-092-437-1

  Query Match             74.5%;  Score 16.4;  DB 3;  Length 963;
  Best Local Similarity   94.4%;  Pred. No. 1.7e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 CCTCTTCAGTCCAATCAA 21
              |||||||||||||| |||
Db        745 CCTCTTCAGTCCAACCAA 728


RESULT 15
US-09-583-110-1197/c
; Sequence 1197, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
;  APPLICANT:  Lynn Doucette-Stamm et al.
;  TITLE OF INVENTION:  Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
;  TITLE OF INVENTION:  Pneumoniae for Diagnostics and Therapeutics
;  FILE REFERENCE:  PATH00-07A
;  CURRENT APPLICATION NUMBER: US/09/583,110
;  CURRENT FILING DATE:  2000-05-26
;  PRIOR APPLICATION NUMBER:  US 09/107,433
;  PRIOR FILING DATE:  1998-06-30
;  PRIOR APPLICATION NUMBER:  US 60/085,131
;  PRIOR FILING DATE:  1998-05-12
;  PRIOR APPLICATION NUMBER:  US 60/051,553
;  PRIOR FILING DATE:  1997-07-02
;  NUMBER OF SEQ ID NOS:  5322
; SEQ ID NO 1197
;   LENGTH: 963
;   TYPE: DNA
;   ORGANISM: Streptococcus pneumoniae
US-09-583-110-1197

  Query Match             74.5%;  Score 16.4;  DB 4;  Length 963;
  Best Local Similarity   94.4%;  Pred. No. 1.7e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 CCTCTTCAGTCCAATCAA 21
              |||||||||||||| |||
Db        745 CCTCTTCAGTCCAACCAA 728


RESULT 16
US-09-107-433-533/c
; Sequence 533, Application US/09107433
; Patent No. 6800744
;   GENERAL INFORMATION:
;        APPLICANT: Lynn A Doucette-Stamm  and David Bush
;        TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
;                            SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE   FOR DIAGNO
;                            THERAPEUTICS
;        NUMBER OF SEQUENCES: 5206
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;             STREET: 100 Beaver Street
;             CITY: Waltham
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02354
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: CD/ROM ISO9660
;             COMPUTER: <Unknown>
;             OPERATING SYSTEM: <Unknown>
;             SOFTWARE: <Unknown>
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/107,433
;             FILING DATE: 30-Jun-1998
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/ 085131
;             FILING DATE: May 12, 1998
;             APPLICATION NUMBER: 60/051553
;             FILING DATE: July 2, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ariniello, Pamela Deneke
;             REGISTRATION NUMBER: 40,489
;             REFERENCE/DOCKET NUMBER: GTC-011
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (781)893-5007
;             TELEFAX: (781)893-8277
;   INFORMATION FOR SEQ ID NO: 533:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 963 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: circular
;        MOLECULE TYPE: DNA (genomic)
;        HYPOTHETICAL: NO
;        ANTI-SENSE: NO
;        ORIGINAL SOURCE:
;             ORGANISM: Streptococcus pneumoniae
;        FEATURE:
;             NAME/KEY:  misc_feature
;             LOCATION: (B) LOCATION 1...963
;        SEQUENCE DESCRIPTION: SEQ ID NO: 533:
US-09-107-433-533

  Query Match             74.5%;  Score 16.4;  DB 4;  Length 963;
  Best Local Similarity   94.4%;  Pred. No. 1.7e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 CCTCTTCAGTCCAATCAA 21
              |||||||||||||| |||
Db        745 CCTCTTCAGTCCAACCAA 728


RESULT 17
US-09-324-258-6/c
; Sequence 6, Application US/09324258
; Patent No. 6723517
; GENERAL INFORMATION:
;  APPLICANT:  Bamdad, Cynthia C.
;  TITLE OF INVENTION:  THE USE OF SELF-ASSEMBLED MONOLAYERS TO
;  TITLE OF INVENTION:  PROBE THE STRUCTURE OF A TARGET MOLECULE
;  FILE REFERENCE:  M1015/7004/TJO
;  CURRENT APPLICATION NUMBER: US/09/324,258
;  CURRENT FILING DATE:  1999-06-02
;  PRIOR APPLICATION NUMBER:  U.S. 60/087,766
;  PRIOR FILING DATE:  1998-06-02
;  NUMBER OF SEQ ID NOS:  20
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 3694
;   TYPE: DNA
;   ORGANISM: Saccharomyces cerevisiae
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (443)...(3088)
;   OTHER INFORMATION: GAL4
US-09-324-258-6
```

```
  Query Match             74.5%;  Score 16.4;  DB 4;  Length 3694;
  Best Local Similarity   94.4%;  Pred. No. 2.2e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 CCTCTTCAGTCCAATCAA 21
              ||||||||| ||||||||
Db        974 CCTCTTCAGACCAATCAA 957


RESULT 18
US-08-961-527-212
; Sequence 212, Application US/08961527
; Patent No. 6420135
;  GENERAL INFORMATION:
;    APPLICANT:  Charles Kunsch
;    TITLE OF INVENTION:  Streptococcus pneumoniae Polynucleotides and Sequences
;    NUMBER OF SEQUENCES:  391
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/961,527
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brookes, A. Anders
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB340P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  212:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3902 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-961-527-212

  Query Match             74.5%;  Score 16.4;  DB 3;  Length 3902;
  Best Local Similarity   94.4%;  Pred. No. 2.2e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 CCTCTTCAGTCCAATCAA 21
              |||||||||||||| |||
Db        449 CCTCTTCAGTCCAACCAA 466


RESULT 19
US-09-949-016-16603
; Sequence 16603, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16603
;   LENGTH: 232547
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-16603

  Query Match             74.5%;  Score 16.4;  DB 4;  Length 232547;
  Best Local Similarity   94.4%;  Pred. No. 4.9e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 CCTCTTCAGTCCAATCAA 21
              ||||||| |||.||||||
Db   190177 CCTCTTCTGTCCAATCAA 190194


RESULT 20
US-09-949-016-145791/c
; Sequence 145791, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 145791
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-145791

  Query Match             73.6%;  Score 16.2;  DB 4;  Length 601;
  Best Local Similarity   85.7%;  Pred. No. 1.9e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 GCCCTCTTCAGTCCAATCAAG 22
              ||   |||||||||||||:|||
Db        391 GCATCCTTCAGTCCAATCAAG 371


RESULT 21
US-09-949-016-146059/c
; Sequence 146059, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
```

```
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 146059
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-146059

  Query Match             73.6%;  Score 16.2;  DB 4;  Length 601;
  Best Local Similarity   85.7%;  Pred. No. 1.9e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 GCCCTCTTCAGTCCAATCAAG 22
              ||   ||||||||||||||||
Db        391 GCATCCTTCAGTCCAATCAAG 371


RESULT 22
US-09-949-016-146327/c
; Sequence 146327, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 146327
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-146327

  Query Match             73.6%;  Score 16.2;  DB 4;  Length 601;
  Best Local Similarity   85.7%;  Pred. No. 1.9e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 GCCCTCTTCAGTCCAATCAAG 22
              ||   ||||||||||||||||
Db        391 GCATCCTTCAGTCCAATCAAG 371


RESULT 23
US-09-949-016-195128/c
; Sequence 195128, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 195128
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-195128

  Query Match             73.6%;  Score 16.2;  DB 4;  Length 601;
  Best Local Similarity   85.7%;  Pred. No. 1.9e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAA 21
              |||||  ||  ||||||||||
Db         64 AGCCCACTCTAGTCCAATCAA 44


RESULT 24
US-09-533-029-9
; Sequence 9, Application US/09533029
; Patent No. 6664446
; GENERAL INFORMATION:
;  APPLICANT: Heard, Jacqueline
;  APPLICANT: Broun, Pierre
;  APPLICANT: Riechmann, Jose-Luis
;  APPLICANT: Keddie, James
;  APPLICANT: Pineda, Omaira
;  APPLICANT: Adam, Luc
;  APPLICANT: Samaha, Raymond
;  APPLICANT: Zhang, James
;  APPLICANT: Yu, Guo-Liang
;  APPLICANT: Ratcliffe, Oliver
;  APPLICANT: Pilgrim, Marsha
;  APPLICANT: Jiang, Cai-Zhong
;  APPLICANT: Reuber, Lynne
;  TITLE OF INVENTION: DISEASE-INDUCED POLYNUCLEOTIDES
;  FILE REFERENCE: MBI-010
;  CURRENT APPLICATION NUMBER: US/09/533,029
;  CURRENT FILING DATE: 2000-03-22
;  EARLIER APPLICATION NUMBER: 60/125,814
;  EARLIER FILING DATE: 1999-03-23
;  NUMBER OF SEQ ID NOS: 121
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
;   LENGTH: 627
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   OTHER INFORMATION: G435
US-09-533-029-9

  Query Match             73.6%;  Score 16.2;  DB 4;  Length 627;
  Best Local Similarity   85.7%;  Pred. No. 2e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 GCCCTCTTCAGTCCAATCAAG 22
              || ||||||||||||| || |
Db        249 GCACTCTTCAGTCCAAGCACG 269


RESULT 25
US-09-640-211A-212
; Sequence 212, Application US/09640211A
; Patent No. 6833446
; GENERAL INFORMATION:
;  APPLICANT: Wood, Marion
;  APPLICANT:  Shenk, Michael A.
;  APPLICANT:  McGrath, Annette
;  APPLICANT:  Glenn, Matthew
;  TITLE OF INVENTION: Compositions and Methods for the
;  TITLE OF INVENTION:  Modification of Gene Transcription
;  FILE REFERENCE: 11000.1021C1U
;  CURRENT APPLICATION NUMBER: US/09/640,211A
;  CURRENT FILING DATE:  2000-08-16
;  NUMBER OF SEQ ID NOS: 2368
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 212
;   LENGTH: 850
```

; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-09-640-211A-212

Query Match 73.6%; Score 16.2; DB 4; Length 850;
Best Local Similarity 85.7%; Pred. No. 2.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy          2 GCCCTCTTCAGTCCAATCAAG 22
              ||| |||||||| | ||||||
Db         32 GCCTTCTTCAGTTCTATCAAG 52
```

Search completed: September 6, 2005, 21:59:08
Job time : 69.9062 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 6, 2005, 16:01:23 ; Search time 198.688 Seconds
(without alignments)
655.473 Million cell updates/sec

Title: US-10-729-421-35
Perfect score: 22
Sequence: 1 agccctcttcagtccaatcaag 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	22	100.0	22	12	ADQ30665	Adq30665 West Nile
	2	22	100.0	24	12	ADN36704	Adn36704 West Nile
	3	22	100.0	24	12	ADN36702	Adn36702 West Nile
	4	22	100.0	24	12	ADN36703	Adn36703 West Nile
	5	22	100.0	51	12	ADN36716	Adn36716 West Nile
	6	22	100.0	51	12	ADN36714	Adn36714 West Nile
	7	22	100.0	51	12	ADN36715	Adn36715 West Nile
	8	22	100.0	69	12	ADN36694	Adn36694 West Nile
c	9	22	100.0	365	6	ABK51710	Abk51710 Partial c
c	10	22	100.0	366	8	ABQ76684	Abq76684 WNVCwt DN
c	11	22	100.0	967	12	ADQ30647	Adq30647 West Nile
c	12	22	100.0	10945	13	ADR32078	Adr32078 Genomic D
c	13	22	100.0	10945	13	ADR67768	Adr67768 West Nile
c	14	22	100.0	10975	12	ADN98022	Adn98022 West Nile
c	15	22	100.0	11029	8	ABZ68481	Abz68481 Nucleotid
c	16	22	100.0	11029	10	ABV74821	Abv74821 West Nile
c	17	22	100.0	11029	12	ADN98023	Adn98023 West Nile
	18	21	95.5	22	12	ADN36705	Adn36705 West Nile
	19	21	95.5	49	12	ADN36717	Adn36717 West Nile
	20	20.4	92.7	51	12	ADN36713	Adn36713 West Nile
	21	20	90.9	24	12	ADN36706	Adn36706 West Nile
	22	20	90.9	51	12	ADN36718	Adn36718 West Nile
	23	19	86.4	24	12	ADN36701	Adn36701 West Nile
c	24	18.8	85.5	10962	12	ADK13681	Adk13681 West Nile
c	25	18.8	85.5	44920	12	ADQ97910	Adq97910 Human can
	26	18	81.8	532	1	AAN90221	Aan90221 Malaria-s
	27	17.8	80.9	247	3	AAC09236	Aac09236 Human sec
c	28	17.8	80.9	348	3	AAA43855	Aaa43855 Human sec
c	29	17.8	80.9	425	10	ADF80038	Adf80038 Leukaemia
c	30	17.8	80.9	447	2	AAV49571	Aav49571 Human lym
c	31	17.8	80.9	616	13	ADS19174	Ads19174 Human C-t
c	32	17.8	80.9	655	2	AAV83108	Aav83108 Human C-t
c	33	17.8	80.9	697	2	AAV49570	Aav49570 Human lym
c	34	17.8	80.9	697	2	AAV54641	Aav54641 Nucleotid
c	35	17.8	80.9	697	6	ABL41987	Abl41987 Nucleotid
c	36	17.8	80.9	701	10	ADC38673	Adc38673 Human cDN
c	37	17.8	80.9	759	6	ABK84720	Abk84720 Human cDN
c	38	17.8	80.9	759	6	ABK48103	Abk48103 Human cDN
c	39	17.8	80.9	759	10	ADD18707	Add18707 Human dis
c	40	17.8	80.9	759	13	ADR06491	Adr06491 Human AIC
c	41	17.8	80.9	759	13	ADP54852	Adp54852 Human PRO
	42	17.8	80.9	935	12	ADQ18422	Adq18422 Human sof
c	43	17.8	80.9	1404	12	ADQ22924	Adq22924 Human sof
c	44	17.8	80.9	1620	13	ADT46681	Adt46681 Bacterial
c	45	17.8	80.9	9375	4	AAK84948	Aak84948 Human imm
c	46	17.8	80.9	10301	4	AAK84949	Aak84949 Human imm
	47	17.8	80.9	110000	2	AAV21209_14	Continuation (15 o
c	48	17.2	78.2	1420	6	ABT08161	Abt08161 XisA reco
c	49	17.2	78.2	1441	6	ABT08163	Abt08163 NLS-XisA
c	50	17.2	78.2	1855	2	AAV27359	Aav27359 Streptoco
c	51	17.2	78.2	1855	6	ABQ84827	Abq84827 S. pneumo
c	52	17.2	78.2	1855	10	ADC45152	Adc45152 S. pneumo
	53	17.2	78.2	3119	5	AAS72586	Aas72586 DNA encod
c	54	17.2	78.2	3789	13	ADR93829	Adr93829 Novel S.
c	55	17.2	78.2	3840	10	ABX05894	Abx05894 S. pneumo
c	56	17.2	78.2	3840	12	ADM91840	Adm91840 S pneumon
	57	17.2	78.2	4841	4	AAK52955	Aak52955 Human pol
	58	17.2	78.2	4880	4	AAK51971	Aak51971 Human pol
	59	17.2	78.2	4898	10	ABZ79896	Abz79896 Human nuc
	60	17.2	78.2	4915	12	ADQ18215	Adq18215 Human sof
	61	17.2	78.2	4915	13	ADR26068	Adr26068 Breast ca
	62	17.2	78.2	5037	12	ADQ22765	Adq22765 Human sof
c	63	17.2	78.2	5309	6	ABT08172	Abt08172 Recombina
c	64	17.2	78.2	16080	6	AAD28651	Aad28651 Human Sal
c	65	17.2	78.2	16535	2	AAV52207	Aav52207 Streptoco
	66	17.2	78.2	110000	10	ABS56454_01	Continuation (2 of
c	67	17.2	78.2	163701	13	ABD33351	Abd33351 Murine ca
c	68	17	77.3	17	6	ACN05481	Acn05481 WNV Amber
	69	17	77.3	17	6	ACN09471	Acn09471 WNV minus
c	70	17	77.3	17	6	ACN04728	Acn04728 WNV DNAzy
	71	17	77.3	17	6	ACN09470	Acn09470 WNV minus
	72	17	77.3	17	6	ACN13599	Acn13599 WNV minus
	73	17	77.3	17	6	ACN12229	Acn12229 WNV minus
	74	17	77.3	17	6	ACN09469	Acn09469 WNV minus
c	75	17	77.3	17	6	ACN01443	Acn01443 WNV Inozy
c	76	17	77.3	17	6	ACN05482	Acn05482 WNV Amber
c	77	16.8	76.4	342	8	ABX99067	Abx99067 Rice endo
c	78	16.8	76.4	440	6	ABQ55371	Abq55371 Human ova
c	79	16.8	76.4	594	12	ACH75379	Ach75379 Human gen
c	80	16.8	76.4	742	2	AAV00437	Aav00437 Clone H90
	81	16.8	76.4	943	5	AAS66880	Aas66880 DNA encod
	82	16.8	76.4	4547	6	AAD28652	Aad28652 Mouse Sal
	83	16.8	76.4	143412	11	ACN44512	Acn44512 Mouse gen
	84	16.4	74.5	518	13	ADQ79152	Adq79152 Novel can
c	85	16.4	74.5	650	12	ADK34072	Adk34072 Yeast let
c	86	16.4	74.5	656	13	ADQ50297	Adq50297 Novel can
c	87	16.4	74.5	960	10	ABX06892	Abx06892 S. pneumo
c	88	16.4	74.5	963	3	AAZ46474	Aaz46474 S. pneumo
c	89	16.4	74.5	963	4	AAS55886	Aas55886 Streptoco
c	90	16.4	74.5	963	4	AAS55543	Aas55543 Streptoco
c	91	16.4	74.5	963	8	ACA49937	Aca49937 Prokaryot
c	92	16.4	74.5	963	13	ADR91898	Adr91898 Novel S.
c	93	16.4	74.5	2643	6	ABN79826	Abn79826 Fungal ZB

c	94	16.4	74.5	2646	2	AAQ61607 Aaq61607 Mutated G
c	95	16.4	74.5	2646	13	ADT47700 Adt47700 Bacterial
c	96	16.4	74.5	2811	12	ADJ92822 Adj92822 Saccharom
	97	16.4	74.5	3189	2	AAV65242 Aav65242 DNA encod
c	98	16.4	74.5	3694	6	ABK86400 Abk86400 Yeast GAL
c	99	16.4	74.5	3694	12	ADN60220 Adn60220 S. cerevi
	100	16.4	74.5	3902	2	AAV52345 Aav52345 Streptoco

ALIGNMENTS

RESULT 1
ADQ30665
ID ADQ30665 standard; DNA; 22 BP.
XX
AC ADQ30665;
XX
DT 23-SEP-2004 (first entry)
XX
DE West Nile Virus capsid gene antisense primer WNVVA2.
XX
KW ss; primer; West Nile Virus; diagnosis.
XX
OS West Nile virus.
XX
PN WO2004055159-A2.
XX
PD 01-JUL-2004.
XX
PF 05-DEC-2003; 2003WO-US038750.
XX
PR 12-DEC-2002; 2002US-0432850P.
PR 20-JUN-2003; 2003US-0480431P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Shyamala V;
XX
DR WPI; 2004-488058/46.
XX
PT New isolated oligonucleotides for accurately diagnosing West Nile virus
PT infection or for capturing, detecting and quantitating West Nile virus in
PT blood samples.
XX
PS Claim 1; SEQ ID NO 35; 56pp; English.
XX
CC The invention relates to an isolated oligonucleotide not more than 60
CC nucleotides in length comprising a nucleotide sequence (S1) of at least
CC 10 contiguous nucleotides from any of the 28 nucleotide sequences (e.g.
CC 20, 21 or 23 bp) given in the specification derived from the West Nile
CC Virus (WNV) genome, a nucleotide sequence (S2) having 90% sequence
CC identity to the nucleotide sequence of (S1), or complements of (S1) and
CC (S2). The oligonucleotide further comprises a detectable label at the 5'-
CC end and/or the 3'-end. The detectable label is a fluorescent label
CC selected from 6-carboxyfluorescein (6-FAM), tetramethyl rhodamine
CC (TAMRA), and 2',4',5',7',-tetrachloro-4-7-dichlorofluorescein (TET). The
CC composition and methods are useful for accurately diagnosing West Nile
CC virus infection or for capturing, detecting and quantitating West Nile
CC virus in biological samples, particularly blood samples. This sequence
CC corresponds to a PCR primer to amplify a fragment of the capsid gene of
CC the WNV genome. The fragment is detected using the oligonucleotides of
CC the invention.
XX
SQ Sequence 22 BP; 6 A; 8 C; 3 G; 5 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 12; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCCCTCTTCAGTCCAATCAAG 22
||||||||||||||||||||||
Db 1 AGCCCTCTTCAGTCCAATCAAG 22

RESULT 2
ADN36704
ID ADN36704 standard; DNA; 24 BP.
XX
AC ADN36704;
XX
DT 15-JUL-2004 (first entry)
XX
DE West Nile virus detection-related oligonucleotide probe SeqID26.
XX
KW hybridisation assay probe; nucleic acid detection;
KW target-complementary sequence; flavivirus; West Nile virus; WNV;
KW RNA virus; infection; meningitis; encephalitis;
KW high throughput screening; probe; ss.
XX
OS West Nile virus.
XX
PN WO2004036190-A2.
XX
PD 29-APR-2004.
XX
PF 10-OCT-2003; 2003WO-US033639.
XX
PR 16-OCT-2002; 2002US-0418891P.
PR 25-NOV-2002; 2002US-0429006P.
PR 24-FEB-2003; 2003US-0449810P.
XX
PA (GENP-) GEN-PROBE INC.
XX
PI Linnen JM, Pollner RB, Wu W, Dennis GG, Darby PM;
XX
DR WPI; 2004-389590/36.
XX
PT New hybridization assay probe comprising target-complementary sequence of
PT bases, useful in detecting flavivirus, e.g. West Nile virus.
XX
PS Claim 78; SEQ ID NO 26; 135pp; English.
XX
CC This invention relates to a novel hybridisation assay probe, for
CC detecting a nucleic acid, which is a probe sequence that comprises a
CC target-complementary sequence of bases, and optionally one or more base
CC sequences that are not complementary to the nucleic acid that is to be
CC detected. The hybridisation assay probes and the kits are useful in
CC detecting and amplifying a target nucleic acid sequence, for example
CC flavivirus like West Nile virus, that may be present in a biological
CC sample. West Nile virus (WNV) is an RNA virus that primarily infects
CC birds and culex mosquitos, with humans and horses serving as incidental
CC hosts. Infection of humans can lead to meningitis or encephalitis. The
CC invention may allow for accurate and efficient high throughput screening.
CC The present sequence is that of an oligonucleotide probe which is related
CC to the invention.
XX
SQ Sequence 24 BP; 7 A; 8 C; 3 G; 6 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 12; Length 24;
Best Local Similarity 100.0%; Pred. No. 0.81;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCCCTCTTCAGTCCAATCAAG 22
||||||||||||||||||||||
Db 3 AGCCCTCTTCAGTCCAATCAAG 24

RESULT 3
ADN36702
ID ADN36702 standard; DNA; 24 BP.
XX
AC ADN36702;
XX
DT 15-JUL-2004 (first entry)
XX

```
DE   West Nile virus detection-related oligonucleotide probe SeqID24.
XX
KW   hybridisation assay probe; nucleic acid detection;
KW   target-complementary sequence; flavivirus; West Nile virus; WNV;
KW   RNA virus; infection; meningitis; encephalitis;
KW   high throughput screening; probe; ss.
XX
OS   West Nile virus.
XX
PN   WO2004036190-A2.
XX
PD   29-APR-2004.
XX
PF   10-OCT-2003; 2003WO-US033639.
XX
PR   16-OCT-2002; 2002US-0418891P.
PR   25-NOV-2002; 2002US-0429006P.
PR   24-FEB-2003; 2003US-0449810P.
XX
PA   (GENP-) GEN-PROBE INC.
XX
PI   Linnen JM,  Pollner RB,  Wu W,  Dennis GG,  Darby PM;
XX
DR   WPI; 2004-389590/36.
XX
PT   New hybridization assay probe comprising target-complementary sequence of
PT   bases, useful in detecting flavivirus, e.g. West Nile virus.
XX
PS   Claim 78; SEQ ID NO 24; 135pp; English.
XX
CC   This invention relates to a novel hybridisation assay probe, for
CC   detecting a nucleic acid, which is a probe sequence that comprises a
CC   target-complementary sequence of bases, and optionally one or more base
CC   sequences that are not complementary to the nucleic acid that is to be
CC   detected. The hybridisation assay probes and the kits are useful in
CC   detecting and amplifying a target nucleic acid sequence, for example
CC   flavivirus like West Nile virus, that may be present in a biological
CC   sample. West Nile virus (WNV) is an RNA virus that primarily infects
CC   birds and culex mosquitos, with humans and horses serving as incidental
CC   hosts. Infection of humans can lead to meningitis or encephalitis. The
CC   invention may allow for accurate and efficient high throughput screening.
CC   The present sequence is that of an oligonucleotide probe which is related
CC   to the invention.
XX
SQ   Sequence 24 BP; 7 A; 8 C; 4 G; 5 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 22;  DB 12;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 0.81;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              |||||||||||||||||||||
Db          1 AGCCCTCTTCAGTCCAATCAAG 22


RESULT 4
ADN36703
ID   ADN36703 standard; DNA; 24 BP.
XX
AC   ADN36703;
XX
DT   15-JUL-2004  (first entry)
XX
DE   West Nile virus detection-related oligonucleotide probe SeqID25.
XX
KW   hybridisation assay probe; nucleic acid detection;
KW   target-complementary sequence; flavivirus; West Nile virus; WNV;
KW   RNA virus; infection; meningitis; encephalitis;
KW   high throughput screening; probe; ss.
XX
OS   West Nile virus.
XX
PN   WO2004036190-A2.
XX
PD   29-APR-2004.
XX
PF   10-OCT-2003; 2003WO-US033639.
XX
PR   16-OCT-2002; 2002US-0418891P.
PR   25-NOV-2002; 2002US-0429006P.
PR   24-FEB-2003; 2003US-0449810P.
XX
PA   (GENP-) GEN-PROBE INC.
XX
PI   Linnen JM,  Pollner RB,  Wu W,  Dennis GG,  Darby PM;
XX
DR   WPI; 2004-389590/36.
XX
PT   New hybridization assay probe comprising target-complementary sequence of
PT   bases, useful in detecting flavivirus, e.g. West Nile virus.
XX
PS   Claim 78; SEQ ID NO 25; 135pp; English.
XX
CC   This invention relates to a novel hybridisation assay probe, for
CC   detecting a nucleic acid, which is a probe sequence that comprises a
CC   target-complementary sequence of bases, and optionally one or more base
CC   sequences that are not complementary to the nucleic acid that is to be
CC   detected. The hybridisation assay probes and the kits are useful in
CC   detecting and amplifying a target nucleic acid sequence, for example
CC   flavivirus like West Nile virus, that may be present in a biological
CC   sample. West Nile virus (WNV) is an RNA virus that primarily infects
CC   birds and culex mosquitos, with humans and horses serving as incidental
CC   hosts. Infection of humans can lead to meningitis or encephalitis. The
CC   invention may allow for accurate and efficient high throughput screening.
CC   The present sequence is that of an oligonucleotide probe which is related
CC   to the invention.
XX
SQ   Sequence 24 BP; 6 A; 8 C; 4 G; 6 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 22;  DB 12;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 0.81;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              |||||||||||||||||||||
Db          2 AGCCCTCTTCAGTCCAATCAAG 23


RESULT 5
ADN36716
ID   ADN36716 standard; DNA; 51 BP.
XX
AC   ADN36716;
XX
DT   15-JUL-2004  (first entry)
XX
DE   West Nile virus detection-related oligonucleotide probe SeqID38.
XX
KW   hybridisation assay probe; nucleic acid detection;
KW   target-complementary sequence; flavivirus; West Nile virus; WNV;
KW   RNA virus; infection; meningitis; encephalitis;
KW   high throughput screening; probe; ss.
XX
OS   West Nile virus.
OS   Enterobacteria phage T7.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1. .27
FT                   /*tag=  a
FT                   /note= "T7 promoter sequence"
FT   misc_feature    28. .51
FT                   /*tag=  b
FT                   /note= "WNV-complimentary sequence"
XX
PN   WO2004036190-A2.
XX
```

```
PD   29-APR-2004.
XX
PF   10-OCT-2003; 2003WO-US033639.
XX
PR   16-OCT-2002; 2002US-0418891P.
PR   25-NOV-2002; 2002US-0429006P.
PR   24-FEB-2003; 2003US-0449810P.
XX
PA   (GENP-) GEN-PROBE INC.
XX
PI   Linnen JM,  Pollner RB,  Wu W,  Dennis GG,  Darby PM;
XX
DR   WPI; 2004-389590/36.
XX
PT   New hybridization assay probe comprising target-complementary sequence of
PT   bases, useful in detecting flavivirus, e.g. West Nile virus.
XX
PS   Disclosure; SEQ ID NO 38; 135pp; English.
XX
CC   This invention relates to a novel hybridisation assay probe, for
CC   detecting a nucleic acid, which is a probe sequence that comprises a
CC   target-complementary sequence of bases, and optionally one or more base
CC   sequences that are not complementary to the nucleic acid that is to be
CC   detected. The hybridisation assay probes and the kits are useful in
CC   detecting and amplifying a target nucleic acid sequence, for example
CC   flavivirus like West Nile virus, that may be present in a biological
CC   sample. West Nile virus (WNV) is an RNA virus that primarily infects
CC   birds and culex mosquitos, with humans and horses serving as incidental
CC   hosts. Infection of humans can lead to meningitis or encephalitis. The
CC   invention may allow for accurate and efficient high throughput screening.
CC   The present sequence is that of an oligonucleotide probe which is related
CC   to the invention.
XX
SQ   Sequence 51 BP; 18 A; 12 C; 8 G; 13 T; 0 U; 0 Other;

  Query Match            100.0%;  Score 22;  DB 12;  Length 51;
  Best Local Similarity   100.0%;  Pred. No. 0.9;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              ||||||||||||||||||||||
Db         30 AGCCCTCTTCAGTCCAATCAAG 51


RESULT 6
ADN36714
ID   ADN36714 standard; DNA; 51 BP.
XX
AC   ADN36714;
XX
DT   15-JUL-2004  (first entry)
XX
DE   West Nile virus detection-related oligonucleotide probe SeqID36.
XX
KW   hybridisation assay probe; nucleic acid detection;
KW   target-complementary sequence; flavivirus; West Nile virus; WNV;
KW   RNA virus; infection; meningitis; encephalitis;
KW   high throughput screening; probe; ss.
XX
OS   West Nile virus.
OS   Enterobacteria phage T7.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1. .27
FT                   /*tag=  a
FT                   /note= "T7 promoter sequence"
FT   misc_feature    28. .51
FT                   /*tag=  b
FT                   /note= "WNV-complimentary sequence"
XX
PN   WO2004036190-A2.
XX
PD   29-APR-2004.
XX
PF   10-OCT-2003; 2003WO-US033639.
XX
PR   16-OCT-2002; 2002US-0418891P.
PR   25-NOV-2002; 2002US-0429006P.
PR   24-FEB-2003; 2003US-0449810P.
XX
PA   (GENP-) GEN-PROBE INC.
XX
PI   Linnen JM,  Pollner RB,  Wu W,  Dennis GG,  Darby PM;
XX
DR   WPI; 2004-389590/36.
XX
PT   New hybridization assay probe comprising target-complementary sequence of
PT   bases, useful in detecting flavivirus, e.g. West Nile virus.
XX
PS   Disclosure; SEQ ID NO 36; 135pp; English.
XX
CC   This invention relates to a novel hybridisation assay probe, for
CC   detecting a nucleic acid, which is a probe sequence that comprises a
CC   target-complementary sequence of bases, and optionally one or more base
CC   sequences that are not complementary to the nucleic acid that is to be
CC   detected. The hybridisation assay probes and the kits are useful in
CC   detecting and amplifying a target nucleic acid sequence, for example
CC   flavivirus like West Nile virus, that may be present in a biological
CC   sample. West Nile virus (WNV) is an RNA virus that primarily infects
CC   birds and culex mosquitos, with humans and horses serving as incidental
CC   hosts. Infection of humans can lead to meningitis or encephalitis. The
CC   invention may allow for accurate and efficient high throughput screening.
CC   The present sequence is that of an oligonucleotide probe which is related
CC   to the invention.
XX
SQ   Sequence 51 BP; 18 A; 12 C; 9 G; 12 T; 0 U; 0 Other;

  Query Match            100.0%;  Score 22;  DB 12;  Length 51;
  Best Local Similarity   100.0%;  Pred. No. 0.9;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              ||||||||||||||||||||||
Db         28 AGCCCTCTTCAGTCCAATCAAG 49


RESULT 7
ADN36715
ID   ADN36715 standard; DNA; 51 BP.
XX
AC   ADN36715;
XX
DT   15-JUL-2004  (first entry)
XX
DE   West Nile virus detection-related oligonucleotide probe SeqID37.
XX
KW   hybridisation assay probe; nucleic acid detection;
KW   target-complementary sequence; flavivirus; West Nile virus; WNV;
KW   RNA virus; infection; meningitis; encephalitis;
KW   high throughput screening; probe; ss.
XX
OS   West Nile virus.
OS   Enterobacteria phage T7.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1. .27
FT                   /*tag=  a
FT                   /note= "T7 promoter sequence"
FT   misc_feature    28. .51
FT                   /*tag=  b
FT                   /note= "WNV-complimentary sequence"
XX
PN   WO2004036190-A2.
XX
PD   29-APR-2004.
XX
```

PF 10-OCT-2003; 2003WO-US033639.
XX
PR 16-OCT-2002; 2002US-0418891P.
PR 25-NOV-2002; 2002US-0429006P.
PR 24-FEB-2003; 2003US-0449810P.
XX
PA (GENP-) GEN-PROBE INC.
XX
PI Linnen JM, Pollner RB, Wu W, Dennis GG, Darby PM;
XX
DR WPI; 2004-389590/36.
XX
PT New hybridization assay probe comprising target-complementary sequence of
PT bases, useful in detecting flavivirus, e.g. West Nile virus.
XX
PS Disclosure; SEQ ID NO 37; 135pp; English.
XX
CC This invention relates to a novel hybridisation assay probe, for
CC detecting a nucleic acid, which is a probe sequence that comprises a
CC target-complementary sequence of bases, and optionally one or more base
CC sequences that are not complementary to the nucleic acid that is to be
CC detected. The hybridisation assay probes and the kits are useful in
CC detecting and amplifying a target nucleic acid sequence, for example
CC flavivirus like West Nile virus, that may be present in a biological
CC sample. West Nile virus (WNV) is an RNA virus that primarily infects
CC birds and culex mosquitos, with humans and horses serving as incidental
CC hosts. Infection of humans can lead to meningitis or encephalitis. The
CC invention may allow for accurate and efficient high throughput screening.
CC The present sequence is that of an oligonucleotide probe which is related
CC to the invention.
XX
SQ Sequence 51 BP; 17 A; 12 C; 9 G; 13 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 12; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.9;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCCCTCTTCAGTCCAATCAAG 22
||||||||||||||||||||||
Db 29 AGCCCTCTTCAGTCCAATCAAG 50

RESULT 8
ADN36694
ID ADN36694 standard; DNA; 69 BP.
XX
AC ADN36694;
XX
DT 15-JUL-2004 (first entry)
XX
DE West Nile virus detection-related oligonucleotide probe SeqID16.
XX
KW hybridisation assay probe; nucleic acid detection;
KW target-complementary sequence; flavivirus; West Nile virus; WNV;
KW RNA virus; infection; meningitis; encephalitis;
KW high throughput screening; probe; ss.
XX
OS West Nile virus.
XX
PN WO2004036190-A2.
XX
PD 29-APR-2004.
XX
PF 10-OCT-2003; 2003WO-US033639.
XX
PR 16-OCT-2002; 2002US-0418891P.
PR 25-NOV-2002; 2002US-0429006P.
PR 24-FEB-2003; 2003US-0449810P.
XX
PA (GENP-) GEN-PROBE INC.
XX
PI Linnen JM, Pollner RB, Wu W, Dennis GG, Darby PM;
XX
DR WPI; 2004-389590/36.
XX
PT New hybridization assay probe comprising target-complementary sequence of
PT bases, useful in detecting flavivirus, e.g. West Nile virus.
XX
PS Claim 68; SEQ ID NO 16; 135pp; English.
XX
CC This invention relates to a novel hybridisation assay probe, for
CC detecting a nucleic acid, which is a probe sequence that comprises a
CC target-complementary sequence of bases, and optionally one or more base
CC sequences that are not complementary to the nucleic acid that is to be
CC detected. The hybridisation assay probes and the kits are useful in
CC detecting and amplifying a target nucleic acid sequence, for example
CC flavivirus like West Nile virus, that may be present in a biological
CC sample. West Nile virus (WNV) is an RNA virus that primarily infects
CC birds and culex mosquitos, with humans and horses serving as incidental
CC hosts. Infection of humans can lead to meningitis or encephalitis. The
CC invention may allow for accurate and efficient high throughput screening.
CC The present sequence is that of an oligonucleotide probe which is related
CC to the invention.
XX
SQ Sequence 69 BP; 18 A; 22 C; 14 G; 15 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 12; Length 69;
Best Local Similarity 100.0%; Pred. No. 0.94;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCCCTCTTCAGTCCAATCAAG 22
||||||||||||||||||||||
Db 5 AGCCCTCTTCAGTCCAATCAAG 26

RESULT 9
ABK51710/c
ID ABK51710 standard; cDNA; 365 BP.
XX
AC ABK51710;
XX
DT 27-AUG-2002 (first entry)
XX
DE Partial cDNA for west nile virus capsid protein.
XX
KW Human; ss; IgE leader sequence; west nile virus capsid protein;
KW RNA secondary structure; free energy; gene therapy; cancer;
KW hyperproliferative disease; autoimmune disease; rheumatoid arthritis;
KW multiple sclerosis; Sjogren's syndrome; sarcoidosis; scleroderma;
KW insulin-dependent diabetes mellitus; autoimmune thyroiditis; psoriasis;
KW reactive arthritis; ankylosing spondylitis; polymyositis; vasculitis;
KW dermatomyositis; Crohn's disease; ulcerative colitis.
XX
OS West Nile virus.
XX
PN WO200229088-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-US031451.
XX
PR 04-OCT-2000; 2000US-0237885P.
XX
PA (UYPE-) UNIV PENNSYLVANIA.
XX
PI Weiner DB, Yang J;
XX
DR WPI; 2002-416682/44.
XX
PT Producing recombinant protein for preparing pharmaceutical compounds to
PT treat, e.g., cancers or autoimmune disorders, comprises predicting
PT secondary structure (SS) of mRNA and modifying DNA to give mRNA with SS
PT having increased free energy.
XX
PS Example 2; Fig 1; 48pp; English.
XX

```
CC   The invention relates to producing (M1) a protein (I) in a recombinant
CC   expression system (II) comprising: (a) predicting the secondary structure
CC   of mRNA; (b) modifying the native heterologous DNA sequence where the
CC   mRNA transcribed from the modified DNA has a secondary structure with
CC   increased free energy; and (c) using the modified DNA in (II) for
CC   production of (I). Also included are (1) an injectable pharmaceutical
CC   composition comprising a nucleic acid molecule that includes a modified
CC   coding sequence (IV) encoding a protein operably linked to regulatory
CC   elements, where (IV) comprises a higher AT or AU content relative to the
CC   AT or AU content of the native coding sequence and further comprising a
CC   pharmaceutical carrier and (2) a recombinant viral vector comprising a
CC   nucleic acid molecule that includes (IV). The method is used for
CC   producing a protein in a recombinant expression system. Use of a nucleic
CC   acid or recombinant viral vector that have modified DNA sequences to
CC   improve protein production can be used in gene therapy and for the
CC   treatment of cancers, hyperproliferative diseases, and autoimmune
CC   diseases such as rheumatoid arthritis, multiple sclerosis, Sjogren's
CC   syndrome, sarcoidosis, insulin-dependent diabetes mellitus, autoimmune
CC   thyroiditis, reactive arthritis, ankylosing spondylitis, scleroderma,
CC   polymyositis, dermatomyositis, psoriasis, vasculitis, Crohn's disease and
CC   ulcerative colitis. The present sequence is a cDNA for West Nile virus
CC   capsid protein. Fusion constructs of modified mRNA for the capsid protein
CC   and human IgE leader sequence are used in an experiment to minimise the
CC   free energy of the capsid protein mRNA. Note:The present sequence is not
CC   shown in the specification but was created using the information in
CC   figure 1 and the sequence appearing as ABK51708
XX
SQ   Sequence 365 BP; 103 A; 80 C; 109 G; 73 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 22;  DB 6;  Length 365;
  Best Local Similarity   100.0%;  Pred. No. 1.2;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              ||||||||||||||||||||||
Db         95 AGCCCTCTTCAGTCCAATCAAG 74


RESULT 10
ABQ76684/c
ID   ABQ76684 standard; DNA; 366 BP.
XX
AC   ABQ76684;
XX
DT   13-MAY-2003  (first entry)
XX
DE   WNVCwt DNA fragment.
XX
KW   Capsid protein; WNVCwt; mRNA secondary structure; cancer;
KW   immunosuppressive; antirheumatic; cytostatic; antiulcer; neuroprotective;
KW   antiarthritic; antidiabetic; antithyroid; antipsoriatic; virucide; gene;
KW   antiparasitic; antiallergic; gene therapy; allergen; multiple sclerosis;
KW   protective immune response; hyperproliferative cell; ulcerative colitis;
KW   hyperproliferative disease; psoriasis; autoimmune disease; psoriasis;
KW   rheumatoid arthritis; Sjogren's syndrome; autoimmune thyroiditis;
KW   insulin dependent diabetes mellitus; Crohn's disease; ds.
XX
OS   West nile virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .366
FT                   /*tag=  a
FT                   /product= "WNVCwt"
FT                   /note= "no start or stop codon given"
XX
PN   US2002123099-A1.
XX
PD   05-SEP-2002.
XX
PF   04-OCT-2001; 2001US-00971806.
XX
PR   04-OCT-2000; 2000US-0237885P.
XX
PA   (WEIN/) WEINER D B.
PA   (YANG/) YANG J.
XX
PI   Weiner DB,  Yang J;
XX
DR   WPI; 2003-066795/06.
DR   P-PSDB; ABG73556.
XX
PT   Producing protein in recombinant expression system involves predicting
PT   secondary structure of RNA encoding a protein and increasing free energy
PT   for the secondary structure by modifying sequence of DNA encoding the
PT   RNA.
XX
PS   Example 2; Fig 1; 25pp; English.
XX
CC   This invention describes a novel method for producing a protein by
CC   translation of mRNA from heterologous DNA sequences. The method involves
CC   predicting the secondary structure of mRNA transcribed from a native
CC   heterologous DNA sequence, modifying the sequence where mRNA transcribed
CC   from the modified DNA sequence has a secondary structure with increased
CC   free energy compared to mRNA transcribed from native DNA and using
CC   modified heterologous DNA for protein production. The products of the
CC   invention have immunosuppressive, antirheumatic, cytostatic, antiulcer,
CC   neuroprotective, antiarthritic, antidiabetic, antithyroid, antipsoriatic,
CC   virucide, antiparasitic and antiallergic activity and can be used for
CC   gene therapy. The method described is useful for producing a protein in a
CC   recombinant expression system, preferably a cell free in vitro
CC   transcription and translation system, an in vitro cell expression system,
CC   a DNA construct used in direct DNA injection, or a recombinant vector for
CC   delivery of DNA to an individual. The products of the invention are
CC   useful for eliciting broad immune responses against a target protein,
CC   i.e. proteins specifically associated with pathogens such as viruses,
CC   parasites, allergens, or the individual's own abnormal cells.
CC   Compositions containing the products of the invention confer a broad
CC   based protective immune response against hyperproliferative cells that
CC   are characteristic in hyperproliferative diseases including all forms of
CC   cancer and psoriasis. Such compositions are also useful for treating
CC   individuals suffering from autoimmune diseases including rheumatoid
CC   arthritis, multiple sclerosis, Sjogren's syndrome, insulin dependent
CC   diabetes mellitus, autoimmune thyroiditis, Crohn's disease, ulcerative
CC   colitis and psoriasis. This sequence encodes the West Nile virus wild-
CC   type capsid protein described as WNVCwt in the disclosure of the
CC   invention
XX
SQ   Sequence 366 BP; 103 A; 81 C; 108 G; 74 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 22;  DB 8;  Length 366;
  Best Local Similarity   100.0%;  Pred. No. 1.2;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              ||||||||||||||||||||||
Db         96 AGCCCTCTTCAGTCCAATCAAG 75


RESULT 11
ADQ30647/c
ID   ADQ30647 standard; DNA; 967 BP.
XX
AC   ADQ30647;
XX
DT   23-SEP-2004  (first entry)
XX
DE   West Nile virus internal diagnosis control sequence.
XX
KW   ss; internal control; West Nile Virus; diagnosis.
XX
OS   West Nile virus.
XX
PN   WO2004055159-A2.
XX
PD   01-JUL-2004.
XX
```

PF 05-DEC-2003; 2003WO-US038750.
XX
PR 12-DEC-2002; 2002US-0432850P.
PR 20-JUN-2003; 2003US-0480431P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Shyamala V;
XX
DR WPI; 2004-488058/46.
XX
PT New isolated oligonucleotides for accurately diagnosing West Nile virus
PT infection or for capturing, detecting and quantitating West Nile virus in
PT blood samples.
XX
PS Claim 27; SEQ ID NO 17; 56pp; English.
XX
CC The invention relates to an isolated oligonucleotide not more than 60
CC nucleotides in length comprising a nucleotide sequence (S1) of at least
CC 10 contiguous nucleotides from any of the 28 nucleotide sequences (e.g.
CC 20, 21 or 23 bp) given in the specification derived from the West Nile
CC Virus (WNV) genome, a nucleotide sequence (S2) having 90% sequence
CC identity to the nucleotide sequence of (S1), or complements of (S1) and
CC (S2). The oligonucleotide further comprises a detectable label at the 5'-
CC end and/or the 3'-end. The detectable label is a fluorescent label
CC selected from 6-carboxyfluorescein (6-FAM), tetramethyl rhodamine
CC (TAMRA), and 2',4',5',7',-tetrachloro-4-7-dichlorofluorescein (TET). The
CC composition and methods are useful for accurately diagnosing West Nile
CC virus infection or for capturing, detecting and quantitating West Nile
CC virus in biological samples, particularly blood samples. This sequence
CC corresponds to an internal control sequence for the detection of WNV
CC sequences using the oligonucleotides of the invention.
XX
SQ Sequence 967 BP; 273 A; 206 C; 272 G; 216 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 12; Length 967;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCCCTCTTCAGTCCAATCAAG 22
||||||||||||||||||||||
Db 197 AGCCCTCTTCAGTCCAATCAAG 176

RESULT 12
ADR32078/c
ID ADR32078 standard; DNA; 10945 BP.
XX
AC ADR32078;
XX
DT 18-NOV-2004 (first entry)
XX
DE Genomic DNA of a West Nile virus.
XX
KW analysis; target; real time PCR; ds; genomic.
XX
OS West Nile virus.
XX
PN WO2004072230-A2.
XX
PD 26-AUG-2004.
XX
PF 10-FEB-2004; 2004WO-US002012.
XX
PR 10-FEB-2003; 2003US-00361004.
XX
PA (CLEA-) CLEARANT INC.
XX
PI Mckenney K, Gillmeister L, Marlowe K, Armistead D;
XX
DR WPI; 2004-625843/60.
XX
PT Analyzing a target nucleic acid sequence in a biological material by real
PT time PCR using nucleic acid primers that are separated by at least 750
PT nucleic acid residues in the target sequence.
XX
PS Disclosure; SEQ ID NO 5; 96pp; English.
XX
CC The invention relates to a novel method for analysing a target nucleic
CC acid sequence in a biological material. The method comprises adding at
CC least two nucleic acid primers that hybridise under stringent conditions
CC to predetermined nucleic acid sequences of the target nucleic acid
CC sequence that are separated by at least 750 nucleic acid residues,
CC amplifying the target nucleic acid sequence by PCR, and detecting and
CC quantifying the target nucleic acid sequence. The methods and
CC compositions of the present invention are useful for analysing a target
CC nucleic acid sequence in a biological material by real time PCR using
CC nucleic acid primers that are separated by at least 750 nucleic acid
CC residues in the target sequence. This polynucleotide sequence represents
CC the genomic DNA of a West Nile virus used in the target analysis method
CC of the invention.
XX
SQ Sequence 10945 BP; 2999 A; 2457 C; 3143 G; 2346 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 13; Length 10945;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCCCTCTTCAGTCCAATCAAG 22
||||||||||||||||||||||
Db 153 AGCCCTCTTCAGTCCAATCAAG 132

RESULT 13
ADR67768/c
ID ADR67768 standard; DNA; 10945 BP.
XX
AC ADR67768;
XX
DT 18-NOV-2004 (first entry)
XX
DE West Nile virus DNA detected by novel detection method.
XX
KW ds; detection; pathogen.
XX
OS West Nile virus.
XX
PN WO2004072231-A2.
XX
PD 26-AUG-2004.
XX
PF 10-FEB-2004; 2004WO-US002013.
XX
PR 10-FEB-2003; 2003US-00361002.
XX
PA (CLEA-) CLEARANT INC.
XX
PI Mckenney K, Gillmeister L, Marlowe K, Armistead D;
XX
DR WPI; 2004-625844/60.
XX
PT Determining level of potentially active biological pathogens in
PT biological material, by adding nucleic acid primer pairs to biological
PT material, amplifying target nucleic acid by PCR, detecting and
PT quantifying target nucleic acid.
XX
PS Disclosure; SEQ ID NO 5; 111pp; English.
XX
CC The invention relates to a method of determining (M1) level of
CC potentially active biological pathogens in biological material, involves
CC adding at least two nucleic acid primer pairs to biological material,
CC amplifying target nucleic acid sequences by PCR, and detecting and
CC quantifying target nucleic acid sequences, where quantity of the nucleic
CC acid sequences is proportional to number of biological pathogens in
CC biological material. (M1) is useful for determining level of potentially
CC active biological pathogens in a biological material such as cells,

```
CC   tissues, blood or blood components, proteins, enzymes, immunoglobulins,
CC   botanicals, food, ligaments, tendons, nerves, bone, teeth, skin grafts,
CC   bone marrow, heart valves, cartilage, corneas, arteries, veins, organs,
CC   lipids, carbohydrates, collagen, chitin and its derivatives, forensic
CC   samples, mummified material, human or animal remains, stem cells, islet
CC   of Langerhans cells, cells for transplantation, red blood cells, white
CC   blood cells or platelets. The biological pathogen is chosen from
CC   bacteria, viruses, fungi and single cell parasites. The biological
CC   pathogen is chosen from Aspergillus , Candida , Histoplasma ,
CC   Saccharomyces , Coccidioides , Cryptococcus , Escherichia , Bacillus ,
CC   Campylobacter , Helicobacter , Listeria , Clostridium , Streptococcus ,
CC   Enterococcus , Staphylococcus , Brucella , Haemophilus , Salmonella ,
CC   Yersinia , Pseudomonas , Serratia , Enterobacter , Klebsiella , Proteus ,
CC   Citrobacter , Corynebacterium , Propionibacterium and Coxiella . The
CC   biological pathogen is chosen from Adeno-associated virus (AAV),
CC   California encephalitis virus, Coronavirus, Coxsackievirus-A,
CC   Coxsackievirus-B, Eastern equine encephalitis Virus (EEEV), Echovirus,
CC   Hantavirus, Hepatitis A virus (HAV), Hepatitis C virus (HCV), Hepatitis
CC   delta virus (HDV), Hepatitis E virus (HEV), Hepatitis G virus (HGV), HIV,
CC   Human T-lymphotrophic virus (HTLV), Influenza virus (Flu virus), Measles
CC   virus (Rubeola), Mumps virus, Norwalk virus, Parainfluenza virus, Polio
CC   virus, Rabies virus, Respiratory Syncytial virus, Rhinovirus, Rubella
CC   virus, Saint Louis encephalitis virus, Western equine encephalitis virus
CC   (WEEV), Yellow fever virus, Adenovirus, Cytomegalovirus (CMV), Epstein-
CC   Barr virus (EBV), Hepatitis B virus (HBV), Herpes simplex virus 1, Herpes
CC   simplex virus 2, Molluscum contagiosum, Papilloma virus (HPV), Smallpox
CC   virus (Variola), Vaccinia virus, Venezuelan equine encephalitis virus
CC   (VEEV), Ebola virus, West Nile virus, Human Parvovirus B19 and Rotavirus.
CC   (M1) is useful for determining the effectiveness of a sterilization
CC   process applied to a biological material. (M1) is useful in determining
CC   whether the biological pathogen is inactive or active. (M1) enables
CC   determination of whether the particular biological pathogen is present in
CC   a biological material as shown by amplification of first target sequence
CC   and whether the biological pathogen is inactive or active. (M1) enables
CC   evaluation of the effectiveness of sterilization processes, and
CC   determination of both the original level and the residual level of
CC   potentially active biological pathogens. This sequence corresponds to a
CC   West Nile virus DNA detected by the method of the invention.
XX
SQ   Sequence 10945 BP; 2999 A; 2457 C; 3143 G; 2346 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 22;  DB 13;  Length 10945;
  Best Local Similarity   100.0%;  Pred. No. 1.9;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              ||||||||||||||||||||||
Db        153 AGCCCTCTTCAGTCCAATCAAG 132


RESULT 14
ADN98022/c
ID   ADN98022 standard; DNA; 10975 BP.
XX
AC   ADN98022;
XX
DT   29-JUL-2004  (first entry)
XX
DE   West Nile Virus isolate 2741 complete genome sequence.
XX
KW   ds; West Nile Virus; envelope protein; glycoprotein E; flavivirus;
KW   Japanese encephalitis virus; Dengue virus; St Louis encephalitis virus.
XX
OS   West Nile virus.
XX
PN   WO2004040263-A2.
XX
PD   13-MAY-2004.
XX
PF   31-OCT-2003; 2003WO-US034823.
XX
PR   31-OCT-2002; 2002US-0422755P.
PR   06-JUN-2003; 2003US-0476513P.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Wong SJ,  Pei-Yong S;
XX
DR   WPI; 2004-400223/37.
DR   GENBANK; AF206518.
XX
PT   New diagnostic kit comprising West Nile Virus (WNV) envelope protein
PT   reactive with antibody against WNV and cross-reactive with antibody
PT   against a flavivirus, useful in diagnosing flavivirus infection caused by
PT   DENV, WNV, JEV or SLEV.
XX
PS   Disclosure; Fig 37; 212pp; English.
XX
CC   The invention relates to a diagnostic kit comprising at least one
CC   isolated and purified polypeptide comprising a West Nile Virus (WNV)
CC   envelope (E) protein or its immunogenic fragment having a native
CC   conformation or non-denatured structure and that is reactive with
CC   antibodies against WNV and cross-reactive with antibodies against a
CC   flavivirus. The diagnostic kit is useful in diagnosing flavivirus
CC   infection caused by DENV, WNV, JEV or SLEV. This sequence corresponds to
CC   the complete nucleotide sequence of the WNV isolate 2741.
XX
SQ   Sequence 10975 BP; 3007 A; 2460 C; 3149 G; 2359 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 22;  DB 12;  Length 10975;
  Best Local Similarity   100.0%;  Pred. No. 2;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              ||||||||||||||||||||||
Db        177 AGCCCTCTTCAGTCCAATCAAG 156


RESULT 15
ABZ68481/c
ID   ABZ68481 standard; DNA; 11029 BP.
XX
AC   ABZ68481;
XX
DT   22-APR-2003  (first entry)
XX
DE   Nucleotide sequence of the genome of West Nile virus IS-98-ST1.
XX
KW   WNV; IS-98-ST1; Flavivirus; infection; encephalitis; gene; ss.
XX
OS   West nile virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             97. .10397
FT                   /*tag=  a
FT                   /product= "polyprotein"
XX
PN   WO200281511-A1.
XX
PD   17-OCT-2002.
XX
PF   04-APR-2002; 2002WO-FR001168.
XX
PR   04-APR-2001; 2001FR-00004599.
PR   06-SEP-2001; 2001FR-00011525.
XX
PA   (INSP ) INST PASTEUR.
PA   (KIMR-) KIMRON VETERINARY INST.
XX
PI   Despres P,  Deubel V,  Guenet J,  Drouet M,  Malkinson M,  Banet C;
PI   Frenkiel M,  Courageot M,  Coulibaly F,  Catteau A,  Flamand M,  Weber P;
PI   Ceccaldi P;
XX
DR   WPI; 2003-058498/05.
DR   P-PSDB; ABP70647.
XX
```

```
PT   New neurovirulent strain of West Nile virus, useful in diagnosis and
PT   screening for antiviral agents, also related nucleic acids, proteins and
PT   antibodies.
XX
PS   Claim 1; Page 34-49; 68pp; French.
XX
CC   The present sequence represents the genome of a strain of West Nile virus
CC   (WNV), designated IS-98-ST1. This strain is a neuroinvasive and
CC   neurovirulent strain of WNV. Polynucleotides and polypeptides derived
CC   from the IS-98-ST1 genome are useful for diagnosis and prognosis of
CC   Flavivirus infection, specifically WNV-mediated encephalitis. They are
CC   also useful to raise specific antibodies, for recombinant expression of
CC   WNV proteins or peptides (for diagnosis, production of antibodies and
CC   identification of specific binding partners in cells), for identifying
CC   cellular genes implicated in resistance to viral infection, and for
CC   screening for anti-Flavivirus agents
XX
SQ   Sequence 11029 BP; 3019 A; 2471 C; 3167 G; 2372 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 22;  DB 8;  Length 11029;
  Best Local Similarity   100.0%;  Pred. No. 2;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              ||||||||||||||||||||||
Db        195 AGCCCTCTTCAGTCCAATCAAG 174


RESULT 16
ABV74821/c
ID   ABV74821 standard; DNA; 11029 BP.
XX
AC   ABV74821;
XX
DT   28-MAR-2003  (first entry)
XX
DE   West Nile virus strain NY99-flamingo 382-99 complete genome.
XX
KW   Virucide; hepatotropic; antiinflammatory; antiviral; OAS;
KW   2'-5'-oligoadenylate synthase; Flavivirus infection; gene; ss.
XX
OS   West Nile Virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             97. .10398
FT                   /*tag=  a
FT                   /product= "West Nile Virus protein"
XX
PN   WO200281741-A2.
XX
PD   17-OCT-2002.
XX
PF   04-APR-2002; 2002WO-FR001169.
XX
PR   04-APR-2001; 2001FR-00004598.
XX
PA   (INSP ) INST PASTEUR.
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Guenet J,  Mashimo T,  Simon-Chazottes D,  Montagutelli X;
PI   Frenkiel M,  Despres P,  Deubel V,  Bonhomme F,  Lucas M;
XX
DR   WPI; 2003-058566/05.
DR   P-PSDB; ABB98821.
XX
PT   Identifying stimulators of oligoadenylate synthase family genes, useful
PT   as antiviral agents against Flavivirus, also mutated genes responsible
PT   for sensitivity to virus.
XX
PS   Example 1; Page 52-67; 93pp; French.
XX
CC   The present invention relates to a method for identifying compounds (I)
CC   that can stimulate a gene of the OAS (2'-5'-oligoadenylate synthase)
CC   family. The method comprises: (a) inducing expression of the OAS gene in
CC   a culture of cells from a non-human mammal (Flvr/Flvr or Flvr/Flvs;
CC   indicating resistance or sensitivity to Flavivirus infection); (b)
CC   treating cells with test compound; and (c) measuring activity of OAS gene
CC   relative to a control. (I) are potentially useful as antiviral agents for
CC   treating infections by Flaviviruses (e.g. hepatitis C; dengue; yellow
CC   fever and various forms of encephalitis). Genomic OAS DNA and derived
CC   cDNA, also the encoded proteins, are useful: (a) for treating Flavivirus
CC   infection; (b) in screening for anti-flavivirus agents, and (c) for
CC   evaluating sensitivity of subjects to Flavivirus infection and their
CC   likely response to interferon treatment, e.g. to identify patients at
CC   risk of developing severe forms of such infections. The present sequence
CC   is West Nile Virus strain NY99-flamingo 382-99 (IS-98-STI) complete
CC   genome, which was used in an example from the invention. West Nile Virus
CC   is one such Flavivirus
XX
SQ   Sequence 11029 BP; 3019 A; 2471 C; 3167 G; 2372 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 22;  DB 10;  Length 11029;
  Best Local Similarity   100.0%;  Pred. No. 2;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              ||||||||||||||||||||||
Db        195 AGCCCTCTTCAGTCCAATCAAG 174


RESULT 17
ADN98023/c
ID   ADN98023 standard; DNA; 11029 BP.
XX
AC   ADN98023;
XX
DT   29-JUL-2004  (first entry)
XX
DE   West Nile Virus isolate 3356 complete genome sequence.
XX
KW   ds; West Nile Virus; envelope protein; glycoprotein E; flavivirus;
KW   Japanese encephalitis virus; Dengue virus; St Louis encephalitis virus.
XX
OS   West Nile virus.
XX
PN   WO2004040263-A2.
XX
PD   13-MAY-2004.
XX
PF   31-OCT-2003; 2003WO-US034823.
XX
PR   31-OCT-2002; 2002US-0422755P.
PR   06-JUN-2003; 2003US-0476513P.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Wong SJ,  Pei-Yong S;
XX
DR   WPI; 2004-400223/37.
DR   GENBANK; AF404756.
XX
PT   New diagnostic kit comprising West Nile Virus (WNV) envelope protein
PT   reactive with antibody against WNV and cross-reactive with antibody
PT   against a flavivirus, useful in diagnosing flavivirus infection caused by
PT   DENV, WNV, JEV or SLEV.
XX
PS   Disclosure; Fig 38; 212pp; English.
XX
CC   The invention relates to a diagnostic kit comprising at least one
CC   isolated and purified polypeptide comprising a West Nile Virus (WNV)
CC   envelope (E) protein or its immunogenic fragment having a native
CC   conformation or non-denatured structure and that is reactive with
CC   antibodies against WNV and cross-reactive with antibodies against a
CC   flavivirus. The diagnostic kit is useful in diagnosing flavivirus
CC   infection caused by DENV, WNV, JEV or SLEV. This sequence corresponds to
CC   the complete nucleotide sequence of the WNV isolate 3356.
```

```
XX
SQ   Sequence 11029 BP; 3017 A; 2466 C; 3172 G; 2374 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 22;  DB 12;  Length 11029;
  Best Local Similarity   100.0%;  Pred. No. 2;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              ||||||||||||||||||||||
Db        195 AGCCCTCTTCAGTCCAATCAAG 174


RESULT 18
ADN36705
ID   ADN36705 standard; DNA; 22 BP.
XX
AC   ADN36705;
XX
DT   15-JUL-2004  (first entry)
XX
DE   West Nile virus detection-related PCR primer SeqID27.
XX
KW   hybridisation assay probe; nucleic acid detection;
KW   target-complementary sequence; flavivirus; West Nile virus; WNV;
KW   RNA virus; infection; meningitis; encephalitis;
KW   high throughput screening; PCR; primer; ss.
XX
OS   West Nile virus.
XX
PN   WO2004036190-A2.
XX
PD   29-APR-2004.
XX
PF   10-OCT-2003; 2003WO-US033639.
XX
PR   16-OCT-2002; 2002US-0418891P.
PR   25-NOV-2002; 2002US-0429006P.
PR   24-FEB-2003; 2003US-0449810P.
XX
PA   (GENP-) GEN-PROBE INC.
XX
PI   Linnen JM,  Pollner RB,  Wu W,  Dennis GG,  Darby PM;
XX
DR   WPI; 2004-389590/36.
XX
PT   New hybridization assay probe comprising target-complementary sequence of
PT   bases, useful in detecting flavivirus, e.g. West Nile virus.
XX
PS   Example 2; SEQ ID NO 27; 135pp; English.
XX
CC   This invention relates to a novel hybridisation assay probe, for
CC   detecting a nucleic acid, which is a probe sequence that comprises a
CC   target-complementary sequence of bases, and optionally one or more base
CC   sequences that are not complementary to the nucleic acid that is to be
CC   detected. The hybridisation assay probes and the kits are useful in
CC   detecting and amplifying a target nucleic acid sequence, for example
CC   flavivirus like West Nile virus, that may be present in a biological
CC   sample. West Nile virus (WNV) is an RNA virus that primarily infects
CC   birds and culex mosquitos, with humans and horses serving as incidental
CC   hosts. Infection of humans can lead to meningitis or encephalitis. The
CC   invention may allow for accurate and efficient high throughput screening.
CC   The present sequence is that of a PCR primer which is related to the
CC   invention.
XX
SQ   Sequence 22 BP; 6 A; 8 C; 2 G; 6 T; 0 U; 0 Other;

  Query Match           95.5%;  Score 21;  DB 12;  Length 22;
  Best Local Similarity   100.0%;  Pred. No. 2.4;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAA 21
              |||||||||||||||||||||
Db          2 AGCCCTCTTCAGTCCAATCAA 22


RESULT 19
ADN36717
ID   ADN36717 standard; DNA; 49 BP.
XX
AC   ADN36717;
XX
DT   15-JUL-2004  (first entry)
XX
DE   West Nile virus detection-related oligonucleotide probe SeqID39.
XX
KW   hybridisation assay probe; nucleic acid detection;
KW   target-complementary sequence; flavivirus; West Nile virus; WNV;
KW   RNA virus; infection; meningitis; encephalitis;
KW   high throughput screening; probe; ss.
XX
OS   West Nile virus.
OS   Enterobacteria phage T7.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1. .27
FT                   /*tag=  a
FT                   /note= "T7 promoter sequence"
FT   misc_feature    28. .49
FT                   /*tag=  b
FT                   /note= "WNV-complimentary sequence"
XX
PN   WO2004036190-A2.
XX
PD   29-APR-2004.
XX
PF   10-OCT-2003; 2003WO-US033639.
XX
PR   16-OCT-2002; 2002US-0418891P.
PR   25-NOV-2002; 2002US-0429006P.
PR   24-FEB-2003; 2003US-0449810P.
XX
PA   (GENP-) GEN-PROBE INC.
XX
PI   Linnen JM,  Pollner RB,  Wu W,  Dennis GG,  Darby PM;
XX
DR   WPI; 2004-389590/36.
XX
PT   New hybridization assay probe comprising target-complementary sequence of
PT   bases, useful in detecting flavivirus, e.g. West Nile virus.
XX
PS   Disclosure; SEQ ID NO 39; 135pp; English.
XX
CC   This invention relates to a novel hybridisation assay probe, for
CC   detecting a nucleic acid, which is a probe sequence that comprises a
CC   target-complementary sequence of bases, and optionally one or more base
CC   sequences that are not complementary to the nucleic acid that is to be
CC   detected. The hybridisation assay probes and the kits are useful in
CC   detecting and amplifying a target nucleic acid sequence, for example
CC   flavivirus like West Nile virus, that may be present in a biological
CC   sample. West Nile virus (WNV) is an RNA virus that primarily infects
CC   birds and culex mosquitos, with humans and horses serving as incidental
CC   hosts. Infection of humans can lead to meningitis or encephalitis. The
CC   invention may allow for accurate and efficient high throughput screening.
CC   The present sequence is that of an oligonucleotide probe which is related
CC   to the invention.
XX
SQ   Sequence 49 BP; 17 A; 12 C; 7 G; 13 T; 0 U; 0 Other;

  Query Match           95.5%;  Score 21;  DB 12;  Length 49;
  Best Local Similarity   100.0%;  Pred. No. 2.7;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAA 21
              |||||||||||||||||||||
Db         29 AGCCCTCTTCAGTCCAATCAA 49
```

```
RESULT 20
ADN36713
ID   ADN36713 standard; DNA; 51 BP.
XX
AC   ADN36713;
XX
DT   15-JUL-2004  (first entry)
XX
DE   West Nile virus detection-related PCR primer SeqID35.
XX
KW   hybridisation assay probe; nucleic acid detection;
KW   target-complementary sequence; flavivirus; West Nile virus; WNV;
KW   RNA virus; infection; meningitis; encephalitis;
KW   high throughput screening; PCR; primer; ss.
XX
OS   West Nile virus.
OS   Enterobacteria phage T7.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1. .27
FT                   /*tag=  a
FT                   /note= "T7 promoter sequence"
FT   misc_feature    28. .51
FT                   /*tag=  b
FT                   /note= "WNV-complimentary sequence"
XX
PN   WO2004036190-A2.
XX
PD   29-APR-2004.
XX
PF   10-OCT-2003; 2003WO-US033639.
XX
PR   16-OCT-2002; 2002US-0418891P.
PR   25-NOV-2002; 2002US-0429006P.
PR   24-FEB-2003; 2003US-0449810P.
XX
PA   (GENP-) GEN-PROBE INC.
XX
PI   Linnen JM,  Pollner RB,  Wu W,  Dennis GG,  Darby PM;
XX
DR   WPI; 2004-389590/36.
XX
PT   New hybridization assay probe comprising target-complementary sequence of
PT   bases, useful in detecting flavivirus, e.g. West Nile virus.
XX
PS   Example 5; SEQ ID NO 35; 135pp; English.
XX
CC   This invention relates to a novel hybridisation assay probe, for
CC   detecting a nucleic acid, which is a probe sequence that comprises a
CC   target-complementary sequence of bases, and optionally one or more base
CC   sequences that are not complementary to the nucleic acid that is to be
CC   detected. The hybridisation assay probes and the kits are useful in
CC   detecting and amplifying a target nucleic acid sequence, for example
CC   flavivirus like West Nile virus, that may be present in a biological
CC   sample. West Nile virus (WNV) is an RNA virus that primarily infects
CC   birds and culex mosquitos, with humans and horses serving as incidental
CC   hosts. Infection of humans can lead to meningitis or encephalitis. The
CC   invention may allow for accurate and efficient high throughput screening.
CC   The present sequence is that of a PCR primer which is related to the
CC   invention.
XX
SQ   Sequence 51 BP; 19 A; 12 C; 8 G; 12 T; 0 U; 0 Other;

  Query Match             92.7%;  Score 20.4;  DB 12;  Length 51;
  Best Local Similarity   95.5%;  Pred. No. 5.3;
  Matches   21;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              || ||||||||||||||||||||
Db         25 AGACCTCTTCAGTCCAATCAAG 46
```

```
RESULT 21
ADN36706
ID   ADN36706 standard; DNA; 24 BP.
XX
AC   ADN36706;
XX
DT   15-JUL-2004  (first entry)
XX
DE   West Nile virus detection-related PCR primer SeqID28.
XX
KW   hybridisation assay probe; nucleic acid detection;
KW   target-complementary sequence; flavivirus; West Nile virus; WNV;
KW   RNA virus; infection; meningitis; encephalitis;
KW   high throughput screening; PCR; primer; ss.
XX
OS   West Nile virus.
XX
PN   WO2004036190-A2.
XX
PD   29-APR-2004.
XX
PF   10-OCT-2003; 2003WO-US033639.
XX
PR   16-OCT-2002; 2002US-0418891P.
PR   25-NOV-2002; 2002US-0429006P.
PR   24-FEB-2003; 2003US-0449810P.
XX
PA   (GENP-) GEN-PROBE INC.
XX
PI   Linnen JM,  Pollner RB,  Wu W,  Dennis GG,  Darby PM;
XX
DR   WPI; 2004-389590/36.
XX
PT   New hybridization assay probe comprising target-complementary sequence of
PT   bases, useful in detecting flavivirus, e.g. West Nile virus.
XX
PS   Claim 78; SEQ ID NO 28; 135pp; English.
XX
CC   This invention relates to a novel hybridisation assay probe, for
CC   detecting a nucleic acid, which is a probe sequence that comprises a
CC   target-complementary sequence of bases, and optionally one or more base
CC   sequences that are not complementary to the nucleic acid that is to be
CC   detected. The hybridisation assay probes and the kits are useful in
CC   detecting and amplifying a target nucleic acid sequence, for example
CC   flavivirus like West Nile virus, that may be present in a biological
CC   sample. West Nile virus (WNV) is an RNA virus that primarily infects
CC   birds and culex mosquitos, with humans and horses serving as incidental
CC   hosts. Infection of humans can lead to meningitis or encephalitis. The
CC   invention may allow for accurate and efficient high throughput screening.
CC   The present sequence is that of a PCR primer which is related to the
CC   invention.
XX
SQ   Sequence 24 BP; 7 A; 9 C; 2 G; 6 T; 0 U; 0 Other;

  Query Match             90.9%;  Score 20;  DB 12;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 7.5;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCA 20
              ||||||||||||||||||||
Db          5 AGCCCTCTTCAGTCCAATCA 24
```

```
RESULT 22
ADN36718
ID   ADN36718 standard; DNA; 51 BP.
XX
AC   ADN36718;
XX
DT   15-JUL-2004  (first entry)
XX
DE   West Nile virus detection-related PCR primer SeqID40.
XX
```

```
KW   hybridisation assay probe; nucleic acid detection;
KW   target-complementary sequence; flavivirus; West Nile virus; WNV;
KW   RNA virus; infection; meningitis; encephalitis;
KW   high throughput screening; PCR; primer; ss.
XX
OS   West Nile virus.
OS   Enterobacteria phage T7.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1. .27
FT                   /*tag=  a
FT                   /note= "T7 promoter sequence"
FT   misc_feature    28. .51
FT                   /*tag=  b
FT                   /note= "WNV-complimentary sequence"
XX
PN   WO2004036190-A2.
XX
PD   29-APR-2004.
XX
PF   10-OCT-2003; 2003WO-US033639.
XX
PR   16-OCT-2002; 2002US-0418891P.
PR   25-NOV-2002; 2002US-0429006P.
PR   24-FEB-2003; 2003US-0449810P.
XX
PA   (GENP-) GEN-PROBE INC.
XX
PI   Linnen JM,  Pollner RB,  Wu W,  Dennis GG,  Darby PM;
XX
DR   WPI; 2004-389590/36.
XX
PT   New hybridization assay probe comprising target-complementary sequence of
PT   bases, useful in detecting flavivirus, e.g. West Nile virus.
XX
PS   Example 6; SEQ ID NO 40; 135pp; English.
XX
CC   This invention relates to a novel hybridisation assay probe, for
CC   detecting a nucleic acid, which is a probe sequence that comprises a
CC   target-complementary sequence of bases, and optionally one or more base
CC   sequences that are not complementary to the nucleic acid that is to be
CC   detected. The hybridisation assay probes and the kits are useful in
CC   detecting and amplifying a target nucleic acid sequence, for example
CC   flavivirus like West Nile virus, that may be present in a biological
CC   sample. West Nile virus (WNV) is an RNA virus that primarily infects
CC   birds and culex mosquitos, with humans and horses serving as incidental
CC   hosts. Infection of humans can lead to meningitis or encephalitis. The
CC   invention may allow for accurate and efficient high throughput screening.
CC   The present sequence is that of a PCR primer which is related to the
CC   invention.
XX
SQ   Sequence 51 BP; 18 A; 13 C; 7 G; 13 T; 0 U; 0 Other;

  Query Match             90.9%;  Score 20;  DB 12;  Length 51;
  Best Local Similarity   100.0%;  Pred. No. 8.3;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCA 20
              ||||||||||||||||||||
Db         32 AGCCCTCTTCAGTCCAATCA 51


RESULT 23
ADN36701
ID   ADN36701 standard; DNA; 24 BP.
XX
AC   ADN36701;
XX
DT   15-JUL-2004  (first entry)
XX
DE   West Nile virus detection-related PCR primer SeqID23.
XX
KW   hybridisation assay probe; nucleic acid detection;
KW   target-complementary sequence; flavivirus; West Nile virus; WNV;
KW   RNA virus; infection; meningitis; encephalitis;
KW   high throughput screening; PCR; primer; ss.
XX
OS   West Nile virus.
XX
PN   WO2004036190-A2.
XX
PD   29-APR-2004.
XX
PF   10-OCT-2003; 2003WO-US033639.
XX
PR   16-OCT-2002; 2002US-0418891P.
PR   25-NOV-2002; 2002US-0429006P.
PR   24-FEB-2003; 2003US-0449810P.
XX
PA   (GENP-) GEN-PROBE INC.
XX
PI   Linnen JM,  Pollner RB,  Wu W,  Dennis GG,  Darby PM;
XX
DR   WPI; 2004-389590/36.
XX
PT   New hybridization assay probe comprising target-complementary sequence of
PT   bases, useful in detecting flavivirus, e.g. West Nile virus.
XX
PS   Example 2; SEQ ID NO 23; 135pp; English.
XX
CC   This invention relates to a novel hybridisation assay probe, for
CC   detecting a nucleic acid, which is a probe sequence that comprises a
CC   target-complementary sequence of bases, and optionally one or more base
CC   sequences that are not complementary to the nucleic acid that is to be
CC   detected. The hybridisation assay probes and the kits are useful in
CC   detecting and amplifying a target nucleic acid sequence, for example
CC   flavivirus like West Nile virus, that may be present in a biological
CC   sample. West Nile virus (WNV) is an RNA virus that primarily infects
CC   birds and culex mosquitos, with humans and horses serving as incidental
CC   hosts. Infection of humans can lead to meningitis or encephalitis. The
CC   invention may allow for accurate and efficient high throughput screening.
CC   The present sequence is that of a PCR primer which is related to the
CC   invention.
XX
SQ   Sequence 24 BP; 8 A; 8 C; 3 G; 5 T; 0 U; 0 Other;

  Query Match             86.4%;  Score 19;  DB 12;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 CCTCTTCAGTCCAATCAAG 22
              |||||||||||||||||||
Db          1 CCTCTTCAGTCCAATCAAG 19


RESULT 24
ADK13681/c
ID   ADK13681 standard; DNA; 10962 BP.
XX
AC   ADK13681;
XX
DT   20-MAY-2004  (first entry)
XX
DE   West Nile Virus DNA sequence, SEQ ID 1.
XX
KW   Virucide; Immunostimulant; flavivirus;
KW   envelope protein domain III polypeptide; envelope protein; gene; ss.
XX
OS   West Nile virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             97. .10389
FT                   /*tag=  a
FT                   /product= "West Nile Virus protein"
XX
PN   WO2004016586-A2.
```

XX
PD 26-FEB-2004.
XX
PF 18-AUG-2003; 2003WO-US025681.
XX
PR 16-AUG-2002; 2002US-0403893P.
PR 06-FEB-2003; 2003US-0445581P.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Barrett A, Beasley D, Holbrook M;
XX
DR WPI; 2004-203756/19.
DR P-PSDB; ADK13682.
XX
PT Diagnosing flavivirus infection by contacting a sample from a human or
PT animal with a flavivirus envelope protein domain III polypeptide, and
PT detecting formation of a immunocomplex between the envelope protein and
PT antibodies in the sample.
XX
PS Disclosure; SEQ ID NO 1; 110pp; English.
XX
CC The present invention relates to a method for screening for a flavivirus
CC in a subject or animal host. The method comprises: contacting a sample
CC from the subject with a composition comprising a flavivirus envelope
CC protein domain III polypeptide (ADK13683-ADK13701) under conditions that
CC permit formation of specific immunocomplex between an antibody in the
CC sample and the envelope protein domain III polypeptide; and detecting
CC whether a specific immunocomplex is formed. The present sequence is the
CC coding sequence for West Nile Virus protein, from which E protein
CC envelope protein domain III polypeptide (ADK13683) is derived.
XX
SQ Sequence 10962 BP; 2997 A; 2497 C; 3100 G; 2368 T; 0 U; 0 Other;

Query Match 85.5%; Score 18.8; DB 12; Length 10962;
Best Local Similarity 90.9%; Pred. No. 68;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AGCCCTCTTCAGTCCAATCAAG 22
||||||||| ||||| ||||||
Db 195 AGCCCTCTTTAGTCCTATCAAG 174

RESULT 25
ADQ97910/c
ID ADQ97910 standard; DNA; 44920 BP.
XX
AC ADQ97910;
XX
DT 07-OCT-2004 (first entry)
XX
DE Human cancer associated sequence HD11-022, SEQ ID 887.
XX
KW Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.
XX
OS Homo sapiens.
XX
PN WO2004060304-A2.
XX
PD 22-JUL-2004.
XX
PF 22-DEC-2003; 2003WO-US041389.
XX
PR 27-DEC-2002; 2002US-00330773.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Malandro MS;
XX
DR WPI; 2004-543781/52.
XX
PT New isolated cancer associated nucleic acids comprising at least 10
PT contiguous nucleotides, useful for diagnosing, preventing and/or treating
PT cancers such as leukemia and lymphoma.
XX
PS Claim 1; SEQ ID NO 887; 199pp; English.
XX
CC The present invention relates to cancer associated sequences (ADQ97025-
CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic formate directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 44920 BP; 13235 A; 9200 C; 9459 G; 13026 T; 0 U; 0 Other;

Query Match 85.5%; Score 18.8; DB 12; Length 44920;
Best Local Similarity 90.9%; Pred. No. 84;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AGCCCTCTTCAGTCCAATCAAG 22
||||||||| || |||||||
Db 42318 AGCCCTCTTCAATCTAATCAAG 42297

Search completed: September 6, 2005, 20:39:28
Job time : 202.688 secs

OM nucleic - nucleic search, using sw model

Run on: September 6, 2005, 16:01:23 ; Search time 770.688 Seconds
(without alignments)
1383.200 Million cell updates/sec

Title: US-10-729-421-35
Perfect score: 22
Sequence: 1 agccctcttcagtccaatcaag 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	22	100.0	1648	14	AF375042	AF375042 West Nile
c	2	22	100.0	1648	14	AF375044	AF375044 West Nile
c	3	22	100.0	1648	14	AF375223	AF375223 West Nile
c	4	22	100.0	2440	14	AF194117	AF194117 West Nile
c	5	22	100.0	10945	14	AF202541	AF202541 West Nile
c	6	22	100.0	10975	14	AF206518	AF206518 West Nile
c	7	22	100.0	10989	14	AY268133	AY268133 West Nile
c	8	22	100.0	10998	14	AY278441	AY278441 West Nile
c	9	22	100.0	11029	6	AX576542	AX576542 Sequence
c	10	22	100.0	11029	6	AX577796	AX577796 Sequence
c	11	22	100.0	11029	14	AB185914	AB185914 West Nile
c	12	22	100.0	11029	14	AB185915	AB185915 West Nile
c	13	22	100.0	11029	14	AB185916	AB185916 West Nile
c	14	22	100.0	11029	14	AB185917	AB185917 West Nile
c	15	22	100.0	11029	14	AF196835	AF196835 West Nile
c	16	22	100.0	11029	14	AF260967	AF260967 West Nile
c	17	22	100.0	11029	14	AF404753	AF404753 West Nile
c	18	22	100.0	11029	14	AF404754	AF404754 West Nile
c	19	22	100.0	11029	14	AF404755	AF404755 West Nile
c	20	22	100.0	11029	14	AF404756	AF404756 West Nile
c	21	22	100.0	11029	14	AF481864	AF481864 West Nile
c	22	22	100.0	11029	14	AF533540	AF533540 West Nile
c	23	22	100.0	11029	14	AY289214	AY289214 West Nile
c	24	22	100.0	11057	14	AY688948	AY688948 West Nile
c	25	21	95.5	1648	14	AF375043	AF375043 West Nile
c	26	21	95.5	1648	14	AF375045	AF375045 West Nile
c	27	21	95.5	2323	14	AF130362	AF130362 West Nile
c	28	21	95.5	2323	14	AF130363	AF130363 West Nile
c	29	21	95.5	10842	14	AY278442	AY278442 West Nile
c	30	21	95.5	10845	14	AY277252	AY277252 West Nile
c	31	21	95.5	10972	14	AF317203	AF317203 West Nile
c	32	21	95.5	10984	14	AY262283	AY262283 West Nile
c	33	21	95.5	10989	14	AY268132	AY268132 West Nile
c	34	21	95.5	11028	14	AY490240	AY490240 West Nile
c	35	21	95.5	11029	14	AF260968	AF260968 West Nile
c	36	21	95.5	11029	14	AF260969	AF260969 West Nile
c	37	21	95.5	11029	14	AF404757	AF404757 West Nile
c	38	19.4	88.2	10664	14	KUNCG	D00246 Kunjin viru
c	39	19.4	88.2	11022	14	AY274504	AY274504 Kunjin vi
c	40	19.4	88.2	11022	14	AY274505	AY274505 Kunjin vi
c	41	19.4	88.2	115637	9	HSO20	AL135750 Homo sapi
c	42	19.4	88.2	120769	9	AL662864	AL662864 Human DNA
c	43	18.8	85.5	240	14	WNF42SAA	M32560 West Nile v
c	44	18.8	85.5	10741	14	AY277251	AY277251 West Nile
c	45	18.8	85.5	10962	14	WNFCG	M12294 West Nile v
	46	18.8	85.5	133888	2	AC010769	AC010769 Homo sapi
	47	18.8	85.5	142560	2	AC021906	AC021906 Homo sapi
	48	18.8	85.5	152037	9	HS167A19	AL031427 Human DNA
	49	18.8	85.5	176827	9	AC022306	AC022306 Homo sapi
	50	18.8	85.5	211550	9	AC073912	AC073912 Homo sapi
c	51	18.4	83.6	203841	2	AC141954	AC141954 Rattus no
	52	18.4	83.6	221153	2	AC115962	AC115962 Mus muscu
	53	18.4	83.6	238643	2	AC133255	AC133255 Rattus no
c	54	18.4	83.6	275209	2	AC097164	AC097164 Rattus no
	55	18	81.8	532	6	E64859	E64859 Malaria-spe
	56	17.8	80.9	247	6	AX897448	AX897448 Sequence
	57	17.8	80.9	247	6	BD032981	BD032981 Sequence
	58	17.8	80.9	380	6	AR424247	AR424247 Sequence
	59	17.8	80.9	380	6	AX984941	AX984941 Sequence
	60	17.8	80.9	380	6	BD119800	BD119800 EST and e
c	61	17.8	80.9	425	6	AX778437	AX778437 Sequence
c	62	17.8	80.9	447	6	BD063596	BD063596 Human pro
c	63	17.8	80.9	487	6	CQ687751	CQ687751 Sequence
c	64	17.8	80.9	547	9	AY142147	AY142147 Homo sapi
	65	17.8	80.9	551	9	BC033359	BC033359 Homo sapi
c	66	17.8	80.9	697	6	BD141462	BD141462 Method of
c	67	17.8	80.9	697	6	E17140	E17140 Human mRNA
c	68	17.8	80.9	697	6	BD063621	BD063621 Human pro
c	69	17.8	80.9	697	9	AB015628	AB015628 Homo sapi
	70	17.8	80.9	699	11	BV180382	BV180382 sqnm10924
c	71	17.8	80.9	717	9	BC005254	BC005254 Homo sapi
c	72	17.8	80.9	756	6	CQ727164	CQ727164 Sequence
c	73	17.8	80.9	759	6	AX394284	AX394284 Sequence
c	74	17.8	80.9	759	9	HSAICL	X96719 H.sapiens m
c	75	17.8	80.9	820	3	AF036902	AF036902 Strongylo
c	76	17.8	80.9	1079	3	AF061750	AF061750 Strongylo
c	77	17.8	80.9	15284	1	U67462	U67462 Methanocald
	78	17.8	80.9	16212	9	HSJ593K13	AL121884 Human DNA
	79	17.8	80.9	74625	9	AL596214	AL596214 Human DNA
	80	17.8	80.9	101016	9	AL355294	AL355294 Human DNA
	81	17.8	80.9	110000	6	AR271569_14	Continuation (15 o
c	82	17.8	80.9	110000	8	CR382127_14	Continuation (15 o
c	83	17.8	80.9	110185	9	AC079795	AC079795 Homo sapi
c	84	17.8	80.9	114428	9	AL844525	AL844525 Human DNA
c	85	17.8	80.9	116061	9	AL732324	AL732324 Human DNA
c	86	17.8	80.9	119430	8	ATF9D16	AL035394 Arabidops
c	87	17.8	80.9	131552	9	AL137063	AL137063 Human DNA
	88	17.8	80.9	138232	2	AC012582	AC012582 Homo sapi
	89	17.8	80.9	145820	9	AC020655	AC020655 Homo sapi
c	90	17.8	80.9	147327	9	AC097455	AC097455 Homo sapi
	91	17.8	80.9	158560	9	AC098969	AC098969 Homo sapi
	92	17.8	80.9	161691	2	AC021385	AC021385 Homo sapi

	93	17.8	80.9	165067	9	AL138927	AL138927 Human DNA
c	94	17.8	80.9	165789	2	AC119054	AC119054 Rattus no
c	95	17.8	80.9	169483	9	AC093829	AC093829 Homo sapi
c	96	17.8	80.9	171462	9	AC117832	AC117832 Homo sapi
	97	17.8	80.9	171518	2	AL589697	AL589697 Homo sapi
c	98	17.8	80.9	171710	2	AC013388	AC013388 Homo sapi
	99	17.8	80.9	173031	9	AL359853	AL359853 Human DNA
	100	17.8	80.9	173169	9	AC068538	AC068538 Homo sapi

ALIGNMENTS

RESULT 1
AF375042/c
LOCUS AF375042 1648 bp mRNA linear VRL 14-MAR-2002
DEFINITION West Nile virus isolate WN_0043 polyprotein mRNA, partial cds.
ACCESSION AF375042
VERSION AF375042.1 GI:19421847
KEYWORDS .
SOURCE West Nile virus
ORGANISM West Nile virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Flavivirus; Japanese encephalitis virus group.
REFERENCE 1 (bases 1 to 1648)
AUTHORS Hindiyeh,M., Shulman,L.M., Mendelson,E., Weiss,L., Grossman,Z. and Bin,H.
TITLE Isolation and characterization of West Nile virus from the blood of viremic patients during the 2000 outbreak in Israel
JOURNAL Emerging Infect. Dis. 7 (4), 748-750 (2001)
MEDLINE 21469825
PUBMED 11585544
REFERENCE 2 (bases 1 to 1648)
AUTHORS Hindiyeh,M., Shulman,L.M., Mendelson,E., Grossman,Z., Weiss,L. and Bin,H.
TITLE Direct Submission
JOURNAL Submitted (30-APR-2001) Central Virology Laboratory, Ministry of Health, Public Health Laboratories, Sheba Medical Center, Tel Hashomer 52621, Israel
FEATURES Location/Qualifiers
source 1. .1648
/organism="West Nile virus"
/mol_type="mRNA"
/isolate="WN_0043"
/specific_host="Homo sapiens"
/db_xref="taxon:11082"
CDS <1. .>1648
/codon_start=3
/product="polyprotein"
/protein_id="AAL87745.1"
/db_xref="GI:19421848"
/translation="SLIGLKRAMLSLIDGKGPIRFVLALLAFFRFTAIAPTRAVLDRW
RGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQKKRGGKTGIAVMIGLIASVGAVTLS
NFQGKVMMTVNATDVTDVITIPTAAGKNLCIVRAMDVGYMCDDTITYECPVLSAGNDP
EDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLTVQTHGESTLANKKGAWMDSTKATRY
LVKTESWILRNPGYALVAAVIGWMLGSNTMQRVVFVVLLLLVAPAYSFNCLGMSNRDF
LEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMVNMEAANLAEVRSYCYLATVSDLS
TKAACPTMGEAHNDKRADPAFVCRQGVVDRGWGNGCGLFGKGSIDTCAKFACSTKAIG
RTILKENIKYEVAIFVHGPTTVESHGNYSTQVGATQAGRFSITPAAPSYTLKLGEYGE
VTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWFMDLNLPWSSAGSTVWRNRETLMEFE
EPHATKQSVIALGSQEGALHQALAGAIPVEFSSNTVKLTSG"
ORIGIN

Query Match 100.0%; Score 22; DB 14; Length 1648;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCCCTCTTCAGTCCAATCAAG 22
||||||||||||||||||||||
Db 26 AGCCCTCTTCAGTCCAATCAAG 5

RESULT 2
AF375044/c
LOCUS AF375044 1648 bp mRNA linear VRL 14-MAR-2002
DEFINITION West Nile virus isolate WN_0247 polyprotein mRNA, partial cds.
ACCESSION AF375044
VERSION AF375044.1 GI:19421851
KEYWORDS .
SOURCE West Nile virus
ORGANISM West Nile virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Flavivirus; Japanese encephalitis virus group.
REFERENCE 1 (bases 1 to 1648)
AUTHORS Hindiyeh,M., Shulman,L.M., Mendelson,E., Weiss,L., Grossman,Z. and Bin,H.
TITLE Isolation and characterization of West Nile virus from the blood of viremic patients during the 2000 outbreak in Israel
JOURNAL Emerging Infect. Dis. 7 (4), 748-750 (2001)
MEDLINE 21469825
PUBMED 11585544
REFERENCE 2 (bases 1 to 1648)
AUTHORS Hindiyeh,M., Shulman,L.M., Mendelson,E., Grossman,Z., Weiss,L. and Bin,H.
TITLE Direct Submission
JOURNAL Submitted (30-APR-2001) Central Virology Laboratory, Ministry of Health, Public Health Laboratories, Sheba Medical Center, Tel Hashomer 52621, Israel
FEATURES Location/Qualifiers
source 1. .1648
/organism="West Nile virus"
/mol_type="mRNA"
/isolate="WN_0247"
/specific_host="Homo sapiens"
/db_xref="taxon:11082"
CDS <1. .>1648
/codon_start=3
/product="polyprotein"
/protein_id="AAL87747.1"
/db_xref="GI:19421852"
/translation="SLIGLKRAMLSLIDGKGPIRFVLALLAFFRFTAIAPTRAVLDRW
RGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQKKRGGKTGIAVMIGLIASVGAVTLS
NFQGKVMMTVNATDVTDVITIPTAAGKNLCIVRAMDVGYMCDDTITYECPVLSAGNDP
EDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLTVQTHGESTLANKKGAWMDSTKATRY
LVKTESWILRNPGYALVAAVIGWMLGSNTMQRVVFVVLLLLVAPAYSFNCLGMSNRDF
LEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLAEVRSYCYLATVSDLS
TKAACPTMGEAHNDKRADPAFVCRQGVVDRGWGNGCGLFGKGSIDTCAKFACSTKAIG
RTILKENIKYEVAIFVHGPTTVESHGNYSTQVGATQAGRFSITPAAPSYTLKLGEYGE
VTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWFMDLNLPWSSAGSTVWRNRETLMEFE
EPHATKQSVIALGSQEGALHQALAGAIPVEFSSNTVKLTSG"
ORIGIN

Query Match 100.0%; Score 22; DB 14; Length 1648;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCCCTCTTCAGTCCAATCAAG 22
||||||||||||||||||||||
Db 26 AGCCCTCTTCAGTCCAATCAAG 5

RESULT 3
AF375223/c
LOCUS AF375223 1648 bp RNA linear VRL 02-DEC-2001
DEFINITION West Nile virus polyprotein gene, partial cds.
ACCESSION AF375223
VERSION AF375223.1 GI:17226060
KEYWORDS .
SOURCE West Nile virus
ORGANISM West Nile virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Flavivirus; Japanese encephalitis virus group.
REFERENCE 1 (bases 1 to 1648)
AUTHORS Banet,C., Brill,A., Samina,I., Yadin,H., Straum,Y., Weisman,J., Pokamonski,S., King,R., Deubel,V. and Malkinson,M.
TITLE Phylogenetic relationships of West Nile viruses isolated in Israel

```
            from 1997 to 2000
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1648)
  AUTHORS   Banet,C., Brill,A., Samina,I., Yadin,H., Straum,Y., Weisman,J.,
            Pokamonski,S., King,R., Deubel,V. and Malkinson,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-2001) Kimron Veterinary Institute, Beit Degan
            50250, Israel
FEATURES             Location/Qualifiers
     source          1. .1648
                     /organism="West Nile virus"
                     /virion
                     /mol_type="genomic RNA"
                     /specific_host="white eyed gull"
                     /db_xref="taxon:11082"
     CDS             <1. .>1648
                     /codon_start=3
                     /product="polyprotein"
                     /protein_id="AAL37596.1"
                     /db_xref="GI:17226061"
                     /translation="SLIGLKRAMLSLIDGKGPIRFVLALLAFFRFTAIAPTRAVLDRW
                     RGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQKKRGGKTGIAVMIGLIASVGAVTLS
                     NFQGKVMMTVNATDVTDVITIPTAAGKNLCIVRAMDVGYMCDDTITYECPVLSAGNDP
                     EDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLTVQTHGESTLANKKGAWMDSTKATRY
                     LVKTESWILRNPGYALVAAVIGWMLGSNTMQRVVFVVLLLLVAPAYSFNCLGMSNRDF
                     LEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLAEVRSYCYLATVSDLS
                     TKAACPTMGEAHNDKRADPAFVCRQGVVDRGWGNGCGLFGKGSIDTCAKFACSTKAIG
                     RTILKENIKYEVAIFVHGPTTVESHGNYSTQVGATQAGRFSITPAAPSYTLKLGEYGE
                     VTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWFMDLNLPWSSAGSTVWRNRETLMEFE
                     EPHATKQSVIALGSQEGALHQALAGAIPVEFSSNTVKLTSG"
ORIGIN

  Query Match          100.0%;  Score 22;  DB 14;  Length 1648;
  Best Local Similarity  100.0%;  Pred. No. 1.5;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              ||||||||||||||||||||||
Db         26 AGCCCTCTTCAGTCCAATCAAG 5


RESULT 4
AF194117/c
LOCUS       AF194117               2440 bp    RNA     linear   VRL 19-JAN-2000
DEFINITION  West Nile virus structural protein precursor, gene, partial cds.
ACCESSION   AF194117
VERSION     AF194117.1  GI:6715269
KEYWORDS    .
SOURCE      West Nile virus
  ORGANISM  West Nile virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE   1  (sites)
  AUTHORS   Lanciotti,R.S., Roehrig,J.T., Deubel,V., Smith,J., Parker,M.,
            Steele,K., Crise,B., Volpe,K.E., Crabtree,M.B., Scherret,J.H.,
            Hall,R.A., MacKenzie,J.S., Cropp,C.B., Panigrahy,B., Ostlund,E.,
            Schmitt,B., Malkinson,M., Banet,C., Weissman,J., Komar,N.,
            Savage,H.M., Stone,W., McNamara,T. and Gubler,D.J.
  TITLE     Origin of the West Nile virus responsible for an outbreak of
            encephalitis in the northeastern United States
  JOURNAL   Science 286 (5448), 2333-2337 (1999)
  MEDLINE   20070288
   PUBMED   10600742
REFERENCE   2  (bases 1 to 2440)
  AUTHORS   Parker,M.D., Crise,B.J., Clayton,J.M. and Smith,J.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1999) Virology Division, U.S. Army Medical
            Research Institute of Infectious Diseases, Bldg. 1425 Fort Detrick,
            Frederick, Maryland 21702, USA
FEATURES             Location/Qualifiers
     source          1. .2440
                     /organism="West Nile virus"
                     /virion
                     /mol_type="genomic RNA"
                     /db_xref="taxon:11082"
     CDS             1. .>2440
                     /codon_start=1
                     /product="structural protein precursor"
                     /protein_id="AAF26360.1"
                     /db_xref="GI:6715270"
                     /translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
                     RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
                     KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
                     AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
                     VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
                     VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
                     MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
                     NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQVG
                     ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
                     MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
                     NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
                     PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
                     INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
                     RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
                     ISRQELRCGSGVFIH"
ORIGIN

  Query Match          100.0%;  Score 22;  DB 14;  Length 2440;
  Best Local Similarity  100.0%;  Pred. No. 1.5;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              ||||||||||||||||||||||
Db         99 AGCCCTCTTCAGTCCAATCAAG 78


RESULT 5
AF202541/c
LOCUS       AF202541               10945 bp    RNA     linear   VRL 16-DEC-1999
DEFINITION  West Nile virus strain HNY1999 polyprotein (C, prM, E, NS1, NS2a,
            NS2b, NS3, NS4a, NS4b, NS5) gene, complete cds.
ACCESSION   AF202541
VERSION     AF202541.1  GI:6581069
KEYWORDS    .
SOURCE      West Nile virus
  ORGANISM  West Nile virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE   1  (bases 1 to 10945)
  AUTHORS   Jia,X.Y., Briese,T., Jordan,I., Rambaut,A., Chi,H.C.,
            Mackenzie,J.S., Hall,R.A., Scherret,J. and Lipkin,W.I.
  TITLE     Genetic analysis of West Nile New York 1999 encephalitis virus
  JOURNAL   Lancet 354 (9194), 1971-1972 (1999)
  MEDLINE   20086017
   PUBMED   10622305
REFERENCE   2  (bases 1 to 10945)
  AUTHORS   Jia,X.Y., Briese,T., Jordan,I. and Lipkin,W.I.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1999) Emerging Diseases Laboratory, Dept.
            Microbiology & Molecular Genetics and Neurology, University of
            California, Irvine, 3101 Gillespie Neuroscience Facility, Irvine,
            CA 92697-4292, USA
FEATURES             Location/Qualifiers
     source          1. .10945
                     /organism="West Nile virus"
                     /mol_type="genomic RNA"
                     /strain="HNY1999"
                     /db_xref="taxon:11082"
                     /country="USA: New York"
                     /note="isolated from total brain RNA (patient NYC99002) by
                     RT-PCR amplification"
     CDS             55. .10356
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="AAF18443.1"
                     /db_xref="GI:6581070"
```

```
                     /translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
                     RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
                     KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
                     AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
                     VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
                     VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
                     MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
                     NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQVG
                     ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
                     MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
                     NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
                     PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
                     INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
                     RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAIN
                     ITRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
                     QMWEAVKDELNTPLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
                     FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVREGNTTECDS
                     KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
                     DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
                     TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
                     DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
                     GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
                     DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
                     LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
                     TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
                     ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMVCLAISAYTPWAILPSVV
                     GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
                     TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
                     VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
                     AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
                     LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
                     NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
                     TEIPDRAWNSGYEWITEYTGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
                     PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
                     SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
                     PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
                     TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
                     LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
                     VFFLLMQRKGIGKIGLGGAVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
                     QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
                     LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
                     GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
                     VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
                     TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
                     VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
                     RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
                     KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
                     CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
                     RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
                     PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
                     EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
                     EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
                     AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
                     YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
                     TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
                     RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
                     DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
                     QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
                     IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
                     DEKYVDYMSSLKRYEDTTLVEDTVL"
mat_peptide          55. .423
                     /product="capsid protein"
                     /note="putative; C"
mat_peptide          424. .924
                     /product="precursor of M protein"
                     /note="putative; prM"
mat_peptide          700. .924
                     /product="membrane protein"
                     /note="putative; M"
mat_peptide          925. .2427
                     /product="envelope protein"
                     /note="putative; E"
mat_peptide          2428. .3483
                     /product="non-structural protein NS1"
                     /note="putative"
mat_peptide          3484. .4176
                     /product="non-structural protein NS2a"
                     /note="putative"
mat_peptide          4177. .4569
                     /product="non-structural protein NS2b"
                     /note="putative"
mat_peptide          4570. .6426
                     /product="non-structural protein NS3"
                     /note="putative"
mat_peptide          6427. .6873
                     /product="non-structural protein NS4a"
                     /note="putative"
mat_peptide          6874. .7638
                     /product="non-structural protein NS4b"
                     /note="putative"
mat_peptide          7639. .10353
                     /product="non-structural protein NS5"
                     /note="putative"
ORIGIN

  Query Match             100.0%;  Score 22;  DB 14;  Length 10945;
  Best Local Similarity   100.0%;  Pred. No. 1.5;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              ||||||||||||||||||||||
Db        153 AGCCCTCTTCAGTCCAATCAAG 132


RESULT 6
AF206518/c
LOCUS       AF206518               10975 bp    DNA     linear   VRL 08-MAY-2000
DEFINITION  West Nile virus isolate 2741, complete genome.
ACCESSION   AF206518
VERSION     AF206518.2  GI:7717200
KEYWORDS    .
SOURCE      West Nile virus
  ORGANISM  West Nile virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE   1  (bases 1 to 10975)
  AUTHORS   Anderson,J.F., Andreadis,T.G., Vossbrinck,C.R., Tirrell,S.,
            Wakem,E.M., French,R.A., Garmendia,A.E. and Van Kruiningen,H.J.
  TITLE     Isolation of West Nile virus from mosquitoes, crows, and a Cooper's
            hawk in Connecticut
  JOURNAL   Science 286 (5448), 2331-2333 (1999)
  MEDLINE   20070287
   PUBMED   10600741
REFERENCE   2  (bases 1 to 10975)
  AUTHORS   Vossbrinck,C.R., Anderson,J.F. and Andreadis,T.G.
  TITLE     Genome Sequence of West Nile Virus from Culex pipiens isolate
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 10975)
  AUTHORS   Anderson,J.F., Andreadis,T.G. and Vossbrinck,C.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-1999) Soil and Water, Connecticut Agricultural
            Experiment Station, 123 Huntington Street, New Haven, CT 06511, USA
REFERENCE   4  (bases 1 to 10975)
  AUTHORS   Anderson,J.F., Andreadis,T.G. and Vossbrinck,C.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAY-2000) Soil and Water, Connecticut Agricultural
            Experiment Station, 123 Huntington Street, New Haven, CT 06511, USA
  REMARK    Sequence update by submitter
COMMENT     On May 8, 2000 this sequence version replaced gi:6636507.
FEATURES             Location/Qualifiers
     source          1. .10975
                     /organism="West Nile virus"
                     /mol_type="genomic DNA"
                     /strain="Connecticut 1999"
                     /isolate="2741"
                     /specific_host="Culex pipiens"
```

```
                     /db_xref="taxon:11082"
                     /country="USA:Connecticut, Greenwich-Stanford town line"
     CDS             79. .10380
                     /codon_start=1
                     /product="polyprotein precursor"
                     /protein_id="AAF20205.2"
                     /db_xref="GI:7717201"
                     /translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
                     RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
                     KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
                     AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
                     VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
                     VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
                     MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
                     NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQVG
                     ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
                     MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
                     NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
                     PCKVPISSVASLNDLTPVGRLVTVNPFVSMATANAKVLIELEPPFGDSYIVVGRGEQQ
                     INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
                     RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
                     ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
                     QMWEAVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
                     FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
                     KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
                     DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
                     TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
                     DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
                     GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
                     DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
                     LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
                     TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
                     ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMVCLAISAYTPWAILPSVV
                     GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
                     TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
                     VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
                     AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
                     LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
                     NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
                     TEIPDRAWNSGYEWITEYTGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
                     PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEARVILGEPSAVTAA
                     SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
                     PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
                     TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
                     LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
                     VFFLLMQRKGIGKIGLGGAVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
                     QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
                     LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALPTLARGFPFVDV
                     GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
                     VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
                     TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
                     VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
                     RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
                     KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
                     CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
                     RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
                     PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
                     EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
                     EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
                     AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
                     YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
                     TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
                     RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
                     DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
                     QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
                     IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
                     DEKYVDYMSSLKRYEDTTLVEDTVL"
ORIGIN

  Query Match          100.0%;  Score 22;  DB 14;  Length 10975;
  Best Local Similarity  100.0%;  Pred. No. 1.5;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

  Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
                ||||||||||||||||||||||
  Db        177 AGCCCTCTTCAGTCCAATCAAG 156


  RESULT 7
  AY268133/c
  LOCUS       AY268133               10989 bp    RNA     linear   VRL 03-NOV-2003
  DEFINITION  West Nile virus strain PaH001 polyprotein (pol) gene, complete cds.
  ACCESSION   AY268133
  VERSION     AY268133.1  GI:33242576
  KEYWORDS    .
  SOURCE      West Nile virus (WNV)
    ORGANISM  West Nile virus
              Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
              Flavivirus; Japanese encephalitis virus group.
  REFERENCE   1  (bases 1 to 10989)
    AUTHORS   Charrel,R.N., Brault,A.C., Gallian,P., Lemasson,J.-J., Murgue,B.,
              Murri,S., Pastorino,B., Zeller,H., de chesse,R., de Micco,P. and de
              Lamballerie,X.
    TITLE     Evolutionary relationship between Old World West Nile virus
              strains. Evidence for viral gene flow between africa, the middle
              east, and europe
    JOURNAL   Virology 315 (2), 381-388 (2003)
    MEDLINE   22949215
     PUBMED   14585341
  REFERENCE   2  (bases 1 to 10989)
    AUTHORS   de Lamballerie,X., Brault,A.C., Gallian,P., Lemasson,J., Murgue,B.,
              Murri,S., Pastorino,B., Zeller,H., Dechesse,R., de Micco,P. and
              Charrel,R.N.
    TITLE     Direct Submission
    JOURNAL   Submitted (03-APR-2003) Virology, Medical University, 27 bd Jean
              Moulin, Marseille 13005, France
  FEATURES             Location/Qualifiers
       source          1. .10989
                       /organism="West Nile virus"
                       /mol_type="genomic RNA"
                       /strain="PaH001"
                       /db_xref="taxon:11082"
       gene            77. .10378
                       /gene="pol"
       CDS             77. .10378
                       /gene="pol"
                       /codon_start=1
                       /product="polyprotein"
                       /protein_id="AAQ00999.1"
                       /db_xref="GI:33242577"
                       /translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
                       RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
                       KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
                       AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
                       VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
                       VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
                       MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
                       NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQIG
                       ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
                       MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGTIPVEFSS
                       NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
                       PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
                       INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
                       RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
                       ISRQELRCGSGVFIHNDVKAWMDRYKYYLETPQGLAKIIQKAHKEGVCGLRSVSRLEH
                       QMWEAVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
                       FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
                       KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
                       DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
                       TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
                       DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
                       GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
                       DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
                       LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
                       TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
                       ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMACLAISAYTPWAILPSVV
```

GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
LAPTRVVAAEMAEALRGLPIRYQTSAVTREHNGNEIVDVMCHATLTHRLMSPHRVPNY
NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
TEIPDRAWNSGYEWITEYIGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
SAAQRRGRVGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
VFFLLMQRKGIGKIGLGGAVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
DEKYVDYMSSLKRYEDTTLVEDTVL"

ORIGIN

Query Match 100.0%; Score 22; DB 14; Length 10989;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCCCTCTTCAGTCCAATCAAG 22
||||||||||||||||||||||
Db 175 AGCCCTCTTCAGTCCAATCAAG 154

RESULT 8
AY278441/c
LOCUS AY278441 10998 bp RNA linear VRL 03-MAY-2003
DEFINITION West Nile virus isolate Ast99-901, complete genome.
ACCESSION AY278441
VERSION AY278441.1 GI:30349729
KEYWORDS .
SOURCE West Nile virus (WNV)
ORGANISM West Nile virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Flavivirus; Japanese encephalitis virus group.
REFERENCE 1 (bases 1 to 10998)
AUTHORS Voronina,A.G., Prilipov,A.G., Kinney,R.M., Samokhvalov,E.I.,
Savage,H.M., Alkhovsky,S.V., Tsychia,R., Sadykova,G.K.,
Shatalov,A.G., Usachev,E.V., Mokhonov,V.V., Butenko,A.M.,
Larichev,V.F., Gubler,D.J. and Lvov,D.K.
TITLE Analysis of a new variants of West Nile virus
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 10998)
AUTHORS Voronina,A.G., Prilipov,A.G., Kinney,R.M., Samokhvalov,E.I.,
Savage,H.M., Alkhovsky,S.V., Tsychia,R., Sadykova,G.K.,
Shatalov,A.G., Usachev,E.V., Mokhonov,V.V., Butenko,A.M.,
Larichev,V.F., Gubler,D.J. and Lvov,D.K.
TITLE Direct Submission
JOURNAL Submitted (17-APR-2003) Molecular Genetic, Ivanovsky Virology
Institute, Gamalei 16, Moscow 123098, Russia
FEATURES Location/Qualifiers
source 1. .10998
/organism="West Nile virus"
/mol_type="genomic RNA"
/isolate="Ast99-901"
/isolation_source="human blood in 1999"
/db_xref="taxon:11082"
/country="Russia: Astrakhan, Volga delta"
CDS 97. .10398
/codon_start=1
/product="polyprotein precursor"
/protein_id="AAP22086.1"
/db_xref="GI:30349730"
/translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAYVCRQGVVDRGWG
NGCGLFGKGSIDTCAKFACSTKATGRTILKENIKYEVAIFVHGPTTVESHGNYSTQIG
ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
QMWEAVKDELNTLLKENGVDLSVVVEKQEGMYKPAPKRLTATTEKLEIGWKAWGKSIL
FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
IRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
LLMVGVGSLIKEKRSAAAKKKGASLLCLALASTGLFNPMILAAGLITCDPNRKRGWPA
TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMACLAVSAYTPWAILPSVV
GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
TEIPDRAWNSGYEWITEYIGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALITLLSVMTMG
VFFLLMQRKGIGKIGLGGTVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVVGPDDVEKLTKGKGPKV
RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
DEKYVDYMSSLKRYEDTTLVEDTVL"

ORIGIN

```
  Query Match           100.0%;  Score 22;  DB 14;  Length 10998;
  Best Local Similarity  100.0%;  Pred. No. 1.5;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              ||||||||||||||||||||||
Db        195 AGCCCTCTTCAGTCCAATCAAG 174


RESULT 9
AX576542/c
LOCUS       AX576542               11029 bp    DNA     linear   PAT 08-JAN-2003
DEFINITION  Sequence 1 from Patent WO02081511.
ACCESSION   AX576542
VERSION     AX576542.1  GI:27646162
KEYWORDS    .
SOURCE      Flavivirus sp.
  ORGANISM  Flavivirus sp.
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus.
REFERENCE   1
  AUTHORS   Despres,P., Deubel,V., Guenet,J.L., Drouet,M.T., Malkinson,M.K.,
            Banet,C.K., Frenkiel,M.P., Courageot,M.P., Coulibaly,F.,
            Catteau,A., Flamand,M., Weber,P. and Ceccaldi,P.E.
  TITLE     Neurovirulent strain of the west nile virus and applications
            thereof
  JOURNAL   Patent: WO 02081511-A 1 17-OCT-2002;
            INSTITUT PASTEUR (FR) ; Kimron Veterinary Institute (IL)
FEATURES             Location/Qualifiers
     source          1. .11029
                     /organism="Flavivirus sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:11093"
     CDS             97. .10398
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD60131.1"
                     /db_xref="GI:27646163"
                     /translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
                     RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
                     KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
                     AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
                     VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
                     VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
                     MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
                     NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQVG
                     ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
                     MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
                     NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
                     PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
                     INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
                     RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
                     ISRQELRCGNGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
                     QMWEAVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
                     FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
                     KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
                     DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
                     TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
                     DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
                     GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
                     DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPRLRCLNLDVYRIL
                     LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
                     TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
                     ESDAEITGSSERVDVRLDDGENFQLMNDPGAPWKIWMLRMVCLAISAYTPWAILPSVV
                     GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
                     TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
                     VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
                     AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
                     LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
                     NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
                     TEIPDRAWNSGYEWITEYTGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
                     PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
                     SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMPDNINMPNGLIAQFYQ
                     PEREKVYTMEGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
                     TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
                     LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
                     VFFLLMQRKGIGKIGLGGAVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
                     QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
                     LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
                     GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
                     VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
                     TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
                     VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHSVSRGTAKLRWLVER
                     RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
                     KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
                     CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
                     RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKINNRIERLRREYSSTWHHDENH
                     PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
                     EKVDTKAPEPPEGAKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
                     EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
                     AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
                     YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
                     TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
                     RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
                     DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
                     QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
                     IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
                     DEKYVDYMSSLKRYEDTTLVEDTVL"
ORIGIN

  Query Match           100.0%;  Score 22;  DB 6;  Length 11029;
  Best Local Similarity  100.0%;  Pred. No. 1.5;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              ||||||||||||||||||||||
Db        195 AGCCCTCTTCAGTCCAATCAAG 174


RESULT 10
AX577796/c
LOCUS       AX577796               11029 bp    DNA     linear   PAT 08-JAN-2003
DEFINITION  Sequence 1 from Patent WO02081741.
ACCESSION   AX577796
VERSION     AX577796.1  GI:27647035
KEYWORDS    .
SOURCE      Flavivirus sp.
  ORGANISM  Flavivirus sp.
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus.
REFERENCE   1
  AUTHORS   Guenet,J.L., Mashimo,T., Simon-Chazottes,D., Montagutelli,X.,
            Frenkiel,M.P., Despres,P., Deubel,V., Bonhomme,F. and Lucas,M.
  TITLE     Use of products of genes of the 2'-5' oligoadenylate synthetase
            family (oas) for screening antiviral agents and for detecting
            responsiveness to flaviviridae infection
  JOURNAL   Patent: WO 02081741-A 1 17-OCT-2002;
            INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE
            SCIENTIFIQUE (CNRS) (FR)
FEATURES             Location/Qualifiers
     source          1. .11029
                     /organism="Flavivirus sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:11093"
     CDS             97. .10398
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD60138.1"
                     /db_xref="GI:27647036"
                     /translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
                     RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
                     KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
                     AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
                     VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
                     VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
                     MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
```

NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQVG
ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
ISRQELRCGNGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
QMWEAVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPRLRCLNLDVYRIL
LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
ESDAEITGSSERVDVRLDDGENFQLMNDPGAPWKIWMLRMVCLAISAYTPWAILPSVV
GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
TEIPDRAWNSGYEWITEYTGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMPDNINMPNGLIAQFYQ
PEREKVYTMEGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
VFFLLMQRKGIGKIGLGGAVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHSVSRGTAKLRWLVER
RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKINNRIERLRREYSSTWHHDENH
PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
EKVDTKAPEPPEGAKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
DEKYVDYMSSLKRYEDTTLVEDTVL"

ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 11029;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCCCTCTTCAGTCCAATCAAG 22
||||||||||||||||||||||
Db 195 AGCCCTCTTCAGTCCAATCAAG 174

RESULT 11
AB185914/c
LOCUS AB185914 11029 bp RNA linear VRL 31-JUL-2004
DEFINITION West Nile virus gene for polyprotein precursor protein, complete cds, isolate: 6-LP.
ACCESSION AB185914
VERSION AB185914.2 GI:50872124
KEYWORDS .
SOURCE West Nile virus (WNV)
ORGANISM West Nile virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Flavivirus; Japanese encephalitis virus group.
REFERENCE 1
AUTHORS Shirato,K., Miyoshi,H., Goto,A., Ako,Y., Ueki,T., Kariwa,H. and Takashima,I.
TITLE Correlation between viral envelope glycosylation and neuroinvasiveness of the New York strain of the West Nile virus
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 11029)
AUTHORS Shirato,K., Kariwa,H. and Takashima,I.
TITLE Direct Submission
JOURNAL Submitted (28-JUL-2004) Kazuya Shirato, Graduate School of Veterinary Medicine, Hokkaido University, Laboratory of Public Health, Depertment of Environmental Veterinary Medicine; Kita-19 Nishi-9, Kita-ku, Sapporo, Hokkaido 060-0818, Japan (E-mail:shirato@vetmed.hokudai.ac.jp, Tel:81-11-706-5213(ex.5213), Fax:81-11-706-5213)
COMMENT On Jul 30, 2004 this sequence version replaced gi:50838778.
FEATURES Location/Qualifiers
source 1. .11029
/organism="West Nile virus"
/mol_type="genomic RNA"
/strain="NY99-6922"
/isolate="6-LP"
/db_xref="taxon:11082"
CDS 97. .10398
/codon_start=1
/product="polyprotein precursor protein"
/protein_id="BAD34488.1"
/db_xref="GI:50838779"
/translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQVG
ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
QMWEAVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMVCLAISAYTPWAILPSVV
GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
TEIPDRAWNSGYEWITEYTGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
VFFLLMQRKGIGKIGLGGAVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA

VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
DEKYVDYMSSLKRYEDTTLVEDTVL"

ORIGIN

Query Match 100.0%; Score 22; DB 14; Length 11029;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCCCTCTTCAGTCCAATCAAG 22
||||||||||||||||||||||
Db 195 AGCCCTCTTCAGTCCAATCAAG 174

RESULT 12
AB185915/c
LOCUS AB185915 11029 bp RNA linear VRL 31-JUL-2004
DEFINITION West Nile virus gene for polyprotein precursor protein, complete cds, isolate: 6-SP.
ACCESSION AB185915
VERSION AB185915.2 GI:50872125
KEYWORDS .
SOURCE West Nile virus (WNV)
ORGANISM West Nile virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Flavivirus; Japanese encephalitis virus group.
REFERENCE 1
AUTHORS Shirato,K., Miyoshi,H., Goto,A., Ako,Y., Ueki,T., Kariwa,H. and Takashima,I.
TITLE Correlation between viral envelope glycosylation and neuroinvasiveness of the New York strain of the West Nile virus
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 11029)
AUTHORS Shirato,K., Kariwa,H. and Takashima,I.
TITLE Direct Submission
JOURNAL Submitted (28-JUL-2004) Kazuya Shirato, Graduate School of Veterinary Medicine, Hokkaido University, Laboratory of Public Health, Depertment of Environmental Veterinary Medicine; Kita-19 Nishi-9, Kita-ku, Sapporo, Hokkaido 060-0818, Japan (E-mail:shirato@vetmed.hokudai.ac.jp, Tel:81-11-706-5213(ex.5213), Fax:81-11-706-5213)
COMMENT On Jul 30, 2004 this sequence version replaced gi:50838780.
FEATURES Location/Qualifiers
source 1. .11029
/organism="West Nile virus"
/mol_type="genomic RNA"
/strain="NY99-6922"
/isolate="6-SP"
/db_xref="taxon:11082"
CDS 97. .10398
/codon_start=1
/product="polyprotein precursor protein"
/protein_id="BAD34489.1"
/db_xref="GI:50838781"
/translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYPTQVG
ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
QMWEAVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMVCLAISAYTPWAILPSVV
GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
TEIPDRAWNSGYEWITEYTGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
VFFLLMQRKGIGKIGLGGAVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
DEKYVDYMSSLKRYEDTTLVEDTVL"

ORIGIN

Query Match 100.0%; Score 22; DB 14; Length 11029;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCCCTCTTCAGTCCAATCAAG 22
||||||||||||||||||||||
Db 195 AGCCCTCTTCAGTCCAATCAAG 174

RESULT 13
AB185916/c
LOCUS AB185916 11029 bp RNA linear VRL 30-JUL-2004
DEFINITION West Nile virus gene for polyprotein precursor protein, complete cds, isolate: B-SP.
ACCESSION AB185916

```
VERSION     AB185916.1  GI:50838782
KEYWORDS    .
SOURCE      West Nile virus (WNV)
  ORGANISM  West Nile virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE   1
  AUTHORS   Shirato,K., Miyoshi,H., Goto,A., Ako,Y., Ueki,T., Kariwa,H. and
            Takashima,I.
  TITLE     Correlation between viral envelope glycosylation and
            neuroinvasiveness of the New York strain of the West Nile virus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 11029)
  AUTHORS   Shirato,K., Kariwa,H. and Takashima,I.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUL-2004) Kazuya Shirato, Graduate School of
            Veterinary Medicine, Hokkaido University, Laboratory of Public
            Health, Depertment of Environmental Veterinary Medicine; Kita-19
            Nishi-9, Kita-ku, Sapporo, Hokkaido 060-0818, Japan
            (E-mail:shirato@vetmed.hokudai.ac.jp, Tel:81-11-706-5213(ex.5213),
            Fax:81-11-706-5213)
FEATURES             Location/Qualifiers
     source          1. .11029
                     /organism="West Nile virus"
                     /mol_type="genomic RNA"
                     /strain="BC787"
                     /isolate="B-SP"
                     /db_xref="taxon:11082"
     CDS             97. .10398
                     /codon_start=1
                     /product="polyprotein precursor protein"
                     /protein_id="BAD34490.1"
                     /db_xref="GI:50838783"
                     /translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
                     RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
                     KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVIMTVNATDVTDVITIPTAAGKNLCIVR
                     AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
                     VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
                     VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
                     MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
                     NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYFTQVG
                     ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
                     MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
                     NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
                     PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
                     INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
                     RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
                     ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
                     QMWEAVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
                     FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
                     KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
                     DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
                     TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
                     DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
                     GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
                     DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
                     LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
                     TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
                     ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMVCLAISAYTPWAILPSVV
                     GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
                     TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
                     VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
                     AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
                     LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
                     NLFVMDETHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
                     TEIPDRAWNSGYEWITEYTGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
                     PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
                     SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
                     PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVPYHDRRWCFDGPR
                     TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
                     LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
                     VFFLLMQRKGIGKIGLGGAVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
                     QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
                     LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
                     GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
                     VVDGIVATDVPELERATPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
                     TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
                     VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
                     RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
                     KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
                     CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
                     RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
                     PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
                     EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
                     EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
                     AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
                     YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
                     TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
                     RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
                     DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
                     QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
                     IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
                     DEKYVDYMSSLKRYEDTTLVEDTVL"
ORIGIN

  Query Match           100.0%;  Score 22;  DB 14;  Length 11029;
  Best Local Similarity  100.0%;  Pred. No. 1.5;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              ||||||||||||||||||||||
Db        195 AGCCCTCTTCAGTCCAATCAAG 174


RESULT 14
AB185917/c
LOCUS       AB185917               11029 bp    RNA     linear   VRL 30-JUL-2004
DEFINITION  West Nile virus gene for polyprotein precursor protein, complete
            cds, isolate: B-LP.
ACCESSION   AB185917
VERSION     AB185917.1  GI:50838784
KEYWORDS    .
SOURCE      West Nile virus (WNV)
  ORGANISM  West Nile virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE   1
  AUTHORS   Shirato,K., Miyoshi,H., Goto,A., Ako,Y., Ueki,T., Kariwa,H. and
            Takashima,I.
  TITLE     Correlation between viral envelope glycosylation and
            neuroinvasiveness of the New York strain of the West Nile virus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 11029)
  AUTHORS   Shirato,K., Kariwa,H. and Takashima,I.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUL-2004) Kazuya Shirato, Graduate School of
            Veterinary Medicine, Hokkaido University, Laboratory of Public
            Health, Depertment of Environmental Veterinary Medicine; Kita-19
            Nishi-9, Kita-ku, Sapporo, Hokkaido 060-0818, Japan
            (E-mail:shirato@vetmed.hokudai.ac.jp, Tel:81-11-706-5213(ex.5213),
            Fax:81-11-706-5213)
FEATURES             Location/Qualifiers
     source          1. .11029
                     /organism="West Nile virus"
                     /mol_type="genomic RNA"
                     /strain="BC787"
                     /isolate="B-LP"
                     /db_xref="taxon:11082"
     CDS             97. .10398
                     /codon_start=1
                     /product="polyprotein precursor protein"
                     /protein_id="BAD34491.1"
                     /db_xref="GI:50838785"
                     /translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
                     RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
                     KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVIMTVNATDVTDVITIPTAAGKNLCIVR
```

AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQVG
ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
QMWEAVKDELNTLLKENGVDFSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMVCLAISAYTPWAILPSVV
GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
TEIPDRAWNSGYEWITEYTGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVPYHDRRWCFDGPR
TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
VFFLLMQRKGIGKIGLGGAVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
VVDGIVATDVPELERATPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
DEKYVDYMSSLKRYEDTTLVEDTVL"

ORIGIN

Query Match 100.0%; Score 22; DB 14; Length 11029;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCCCTCTTCAGTCCAATCAAG 22
||||||||||||||||||||||
Db 195 AGCCCTCTTCAGTCCAATCAAG 174

RESULT 15
AF196835/c
LOCUS AF196835 11029 bp RNA linear VRL 07-DEC-2000
DEFINITION West Nile virus strain NY99-flamingo382-99, complete genome.
ACCESSION AF196835
VERSION AF196835.2 GI:11597239
KEYWORDS .
SOURCE West Nile virus
ORGANISM West Nile virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Flavivirus; Japanese encephalitis virus group.
REFERENCE 1 (bases 1 to 11029)
AUTHORS Lanciotti,R.S., Roehrig,J.T., Deubel,V., Smith,J., Parker,M.,
Steele,K., Crise,B., Volpe,K.E., Crabtree,M.B., Scherret,J.H.,
Hall,R.A., MacKenzie,J.S., Cropp,C.B., Panigrahy,B., Ostlund,E.,
Schmitt,B., Malkinson,M., Banet,C., Weissman,J., Komar,N.,
Savage,H.M., Stone,W., McNamara,T. and Gubler,D.J.
TITLE Origin of the West Nile virus responsible for an outbreak of
encephalitis in the northeastern United States
JOURNAL Science 286 (5448), 2333-2337 (1999)
MEDLINE 20070288
PUBMED 10600742
REFERENCE 2 (bases 1 to 11029)
AUTHORS Lanciotti,R., Roehrig,J., Volpe,K. and Panigrahy,B.
TITLE Direct Submission
JOURNAL Submitted (20-OCT-1999) Division of Vector-Borne Diseases, Centers
for Disease Control and Prevention, Rampart Road, Fort Collins, CO
80521, USA
REFERENCE 3 (bases 1 to 11029)
AUTHORS Lanciotti,R., Roehrig,J., Volpe,K. and Panigrahy,B.
TITLE Direct Submission
JOURNAL Submitted (07-DEC-2000) Division of Vector-Borne Diseases, Centers
for Disease Control and Prevention, Rampart Road, Fort Collins, CO
80521, USA
REMARK Sequence update by submitter
COMMENT On Dec 7, 2000 this sequence version replaced gi:6636174.
FEATURES Location/Qualifiers
source 1. .11029
/organism="West Nile virus"
/mol_type="genomic RNA"
/strain="NY99-flamingo382-99"
/db_xref="taxon:11082"
/country="USA: New York"
/note="1999; passage level 0 in chicken embryo"
CDS 97. .10398
/codon_start=1
/product="polyprotein precursor"
/protein_id="AAF20092.2"
/db_xref="GI:11597240"
/translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQVG
ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
QMWEAVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMVCLAISAYTPWAILPSVV
GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS

```
                     AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
                     LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
                     NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
                     TEIPDRAWNSGYEWITEYTGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
                     PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
                     SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
                     PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
                     TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
                     LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
                     VFFLLMQRKGIGKIGLGGAVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
                     QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
                     LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
                     GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
                     VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
                     TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
                     VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
                     RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
                     KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
                     CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
                     RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
                     PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
                     EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
                     EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
                     AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
                     YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
                     TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
                     RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
                     DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
                     QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
                     IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
                     DEKYVDYMSSLKRYEDTTLVEDTVL"
     mat_peptide     97. .465
                     /product="nucleocapsid protein"
     mat_peptide     466. .741
                     /product="pre-membrane protein M"
     mat_peptide     742. .966
                     /product="membrane protein M"
     mat_peptide     967. .2469
                     /product="envelope glycoprotein E"
     mat_peptide     2470. .3525
                     /product="non-structural protein 1"
     mat_peptide     3526. .4218
                     /product="non-structural protein 2A"
     mat_peptide     4219. .4611
                     /product="non-structural protein 2B"
     mat_peptide     4612. .6468
                     /product="non-structural protein 3"
     mat_peptide     6469. .6915
                     /product="non-structural protein 4A"
     mat_peptide     6916. .7680
                     /product="non-structural protein 4B"
     mat_peptide     7681. .10395
                     /product="non-structural protein 5"
ORIGIN

  Query Match             100.0%;  Score 22;  DB 14;  Length 11029;
  Best Local Similarity   100.0%;  Pred. No. 1.5;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              ||||||||||||||||||||||
Db        195 AGCCCTCTTCAGTCCAATCAAG 174


RESULT 16
AF260967/c
LOCUS       AF260967               11029 bp    RNA     linear   VRL 27-AUG-2000
DEFINITION  West Nile virus strain NY99-eqhs, complete genome.
ACCESSION   AF260967
VERSION     AF260967.1  GI:9930133
KEYWORDS    .
SOURCE      West Nile virus
  ORGANISM  West Nile virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE   1  (bases 1 to 11029)
  AUTHORS   Bowen,M., Meyer,R.F., McKinney,N., Morrill,W. and Lanciotti,R.
  TITLE     Complete genomic sequence of West Nile virus equine isolate New
            York 1999
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 11029)
  AUTHORS   Bowen,M., Meyer,R.F., McKinney,N., Morrill,W. and Lanciotti,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2000) Arbovirus Diseases Branch, Centers for
            Disease Control & Prevention, Rampart Road, Fort Collins, CO 80521,
            USA
FEATURES             Location/Qualifiers
     source          1. .11029
                     /organism="West Nile virus"
                     /mol_type="genomic RNA"
                     /strain="NY99-eqhs"
                     /specific_host="horse"
                     /db_xref="taxon:11082"
                     /country="USA: New York"
                     /note="1999"
     CDS             97. .10398
                     /codon_start=1
                     /product="polyprotein precursor"
                     /protein_id="AAG02038.1"
                     /db_xref="GI:9930134"
                     /translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
                     RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
                     KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
                     AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
                     VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
                     VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
                     MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
                     NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQVG
                     ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
                     MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
                     NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
                     PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
                     INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
                     RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
                     ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
                     QMWEAVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
                     FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
                     KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
                     DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
                     TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
                     DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
                     GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAAFFQMAYH
                     DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
                     LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
                     TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
                     ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMVCLAISAYTPWAILPSVV
                     GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
                     TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
                     VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
                     AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
                     LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
                     NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
                     TEIPDRAWNSGYEWITEYTGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
                     PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
                     SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
                     PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVPYHDRRWCFDGPR
                     TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
                     LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
                     VFFLLMQRKGIGKIGLGGAVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
                     QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
                     LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
                     GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
                     VVDGIVATDVPELERATPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
                     TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
                     VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
                     RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
```

```
                     KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
                     CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
                     RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
                     PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
                     EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
                     EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
                     AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
                     YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
                     TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
                     RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
                     DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
                     QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
                     IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
                     DEKYVDYMSSLKRYEDTTLVEDTVL"
     mat_peptide     97. .465
                     /product="nucleocapsid protein C"
     mat_peptide     466. .741
                     /product="pre-membrane protein prM"
     mat_peptide     742. .966
                     /product="membrane protein M"
     mat_peptide     967. .2469
                     /product="envelope glycoprotein E"
     mat_peptide     2470. .3525
                     /product="non-structural protein 1 NS1"
     mat_peptide     3526. .4218
                     /product="non-structural protein 2A NS2A"
     mat_peptide     4219. .4611
                     /product="non-structural protein 2B NS2B"
     mat_peptide     4612. .6468
                     /product="non-structural protein 3 NS3"
     mat_peptide     6469. .6915
                     /product="non-structural protein 4A NS4A"
     mat_peptide     6916. .7680
                     /product="non-structural protein NS4B"
     mat_peptide     7681. .10395
                     /product="non-structural protein NS5"
ORIGIN

  Query Match             100.0%;  Score 22;  DB 14;  Length 11029;
  Best Local Similarity   100.0%;  Pred. No. 1.5;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              ||||||||||||||||||||||
Db        195 AGCCCTCTTCAGTCCAATCAAG 174


RESULT 17
AF404753/c
LOCUS       AF404753               11029 bp ss-RNA     linear   VRL 23-JUL-2002
DEFINITION  West Nile virus isolate WN MD 2000-crow265, complete genome.
ACCESSION   AF404753
VERSION     AF404753.1  GI:21929232
KEYWORDS    .
SOURCE      West Nile virus
  ORGANISM  West Nile virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE   1  (bases 1 to 11029)
  AUTHORS   Lanciotti,R.S., Ebel,G.D., Deubel,V., Kerst,A.J., Murri,S.,
            Meyer,R., Bowen,M., McKinney,N., Morrill,W.E., Crabtree,M.B.,
            Kramer,L.D. and Roehrig,J.T.
  TITLE     Complete genome sequences and phylogenetic analysis of West Nile
            virus strains isolated from the United States, Europe, and the
            Middle East
  JOURNAL   Virology 298 (1), 96-105 (2002)
  MEDLINE   22089180
   PUBMED   12093177
REFERENCE   2  (bases 1 to 11029)
  AUTHORS   Lanciotti,R.S., Ebel,G.D. and Kerst,A.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Division of Vector-Borne Infectious
            Diseases, Centers for Disease Control & Prevention, Rampart Road,
            Fort Collins, CO 80521, USA
FEATURES             Location/Qualifiers
     source          1. .11029
                     /organism="West Nile virus"
                     /mol_type="genomic RNA"
                     /isolate="WN MD 2000-crow265"
                     /specific_host="crow"
                     /db_xref="taxon:11082"
                     /country="USA: Maryland"
     CDS             97. .10398
                     /note="contains capsid, pre-membrane, envelope, NS1, NS2a,
                     NS2b, NS3, NS4a, NS4b, and NS5"
                     /codon_start=1
                     /product="polyprotein precursor"
                     /protein_id="AAM81749.1"
                     /db_xref="GI:21929233"
                     /translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
                     RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
                     KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
                     AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
                     VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
                     VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
                     MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
                     NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQVG
                     ATQAGRFSITPAAPSHTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
                     MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
                     NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
                     PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
                     INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
                     RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
                     ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
                     QMWEAVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
                     FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
                     KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
                     DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
                     TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
                     DMIDPFQLGLLVVFLATQEVLRKRWTAKISLPAILIALLVLVFGGITYTDVLRYVILV
                     GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
                     DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
                     LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
                     TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
                     ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMVCLAISAYTPWAILPSVV
                     GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
                     TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
                     VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
                     AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
                     LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
                     NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
                     TEIPDRAWNSGYEWITEYTGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
                     PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
                     SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
                     PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
                     TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
                     LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
                     VFFLLMQRKGIGKIGLGGAVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
                     QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
                     LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
                     GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
                     VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
                     TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
                     VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
                     RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
                     KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
                     CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
                     RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
                     PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
                     EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
                     EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
                     AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
                     YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
                     TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
                     RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
                     DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
```

```
                    QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
                    IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
                    DEKYVDYMSSLKRYEDTTLVEDTVL"
ORIGIN

  Query Match           100.0%;  Score 22;  DB 14;  Length 11029;
  Best Local Similarity  100.0%;  Pred. No. 1.5;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              ||||||||||||||||||||||
Db        195 AGCCCTCTTCAGTCCAATCAAG 174


RESULT 18
AF404754/c
LOCUS       AF404754               11029 bp ss-RNA     linear   VRL 23-JUL-2002
DEFINITION  West Nile virus isolate WN NJ 2000 MQ5488, complete genome.
ACCESSION   AF404754
VERSION     AF404754.1  GI:21929234
KEYWORDS    .
SOURCE      West Nile virus
  ORGANISM  West Nile virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE   1  (bases 1 to 11029)
  AUTHORS   Lanciotti,R.S., Ebel,G.D., Deubel,V., Kerst,A.J., Murri,S.,
            Meyer,R., Bowen,M., McKinney,N., Morrill,W.E., Crabtree,M.B.,
            Kramer,L.D. and Roehrig,J.T.
  TITLE     Complete genome sequences and phylogenetic analysis of West Nile
            virus strains isolated from the United States, Europe, and the
            Middle East
  JOURNAL   Virology 298 (1), 96-105 (2002)
  MEDLINE   22089180
   PUBMED   12093177
REFERENCE   2  (bases 1 to 11029)
  AUTHORS   Lanciotti,R.S., Ebel,G.D. and Kerst,A.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Division of Vector-Borne Infectious
            Diseases, Centers for Disease Control & Prevention, Rampart Road,
            Fort Collins, CO 80521, USA
FEATURES             Location/Qualifiers
     source          1. .11029
                     /organism="West Nile virus"
                     /mol_type="genomic RNA"
                     /isolate="WN NJ 2000 MQ5488"
                     /specific_host="Culex pipiens pool"
                     /db_xref="taxon:11082"
                     /country="USA: New Jersey"
     CDS             97. .10398
                     /note="contains capsid, pre-membrane, envelope, NS1, NS2a,
                     NS2b, NS3, NS4a, NS4b, and NS5"
                     /codon_start=1
                     /product="polyprotein precursor"
                     /protein_id="AAM81750.1"
                     /db_xref="GI:21929235"
                     /translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
                     RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKE
                     KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
                     AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
                     VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
                     VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
                     MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
                     NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQVG
                     ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
                     MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
                     NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
                     PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
                     INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
                     RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
                     ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
                     QMWEAVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
                     FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
                     KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
                     DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
                     TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
                     DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
                     GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
                     DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
                     LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
                     TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
                     ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMVCLAISAYTPWAILPSVV
                     GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
                     TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
                     VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
                     AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
                     LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
                     NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
                     TEIPDRAWNSGYEWITEYTGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
                     PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
                     SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
                     PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
                     TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
                     LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
                     VFFLLMQRKGIGKIGLGGAVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
                     QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
                     LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
                     GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
                     VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
                     TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
                     VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
                     RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
                     KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
                     CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
                     RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
                     PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
                     EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
                     EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
                     AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
                     YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
                     TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
                     RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
                     DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
                     QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
                     IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
                     DEKYVDYMSSLKRYEDTTLVEDTVL"
ORIGIN

  Query Match           100.0%;  Score 22;  DB 14;  Length 11029;
  Best Local Similarity  100.0%;  Pred. No. 1.5;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              ||||||||||||||||||||||
Db        195 AGCCCTCTTCAGTCCAATCAAG 174


RESULT 19
AF404755/c
LOCUS       AF404755               11029 bp ss-RNA     linear   VRL 23-JUL-2002
DEFINITION  West Nile virus isolate WN NY 2000-grouse3282, complete genome.
ACCESSION   AF404755
VERSION     AF404755.1  GI:21929236
KEYWORDS    .
SOURCE      West Nile virus
  ORGANISM  West Nile virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE   1  (bases 1 to 11029)
  AUTHORS   Lanciotti,R.S., Ebel,G.D., Deubel,V., Kerst,A.J., Murri,S.,
            Meyer,R., Bowen,M., McKinney,N., Morrill,W.E., Crabtree,M.B.,
            Kramer,L.D. and Roehrig,J.T.
  TITLE     Complete genome sequences and phylogenetic analysis of West Nile
            virus strains isolated from the United States, Europe, and the
            Middle East
  JOURNAL   Virology 298 (1), 96-105 (2002)
```

```
  MEDLINE   22089180
   PUBMED   12093177
REFERENCE   2  (bases 1 to 11029)
  AUTHORS   Ebel,G.D., Kerst,A.J. and Lanciotti,R.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Division of Vector-Borne Infectious
            Diseases, Centers for Disease Control & Prevention, Rampart Road,
            Fort Collins, CO 80521, USA
FEATURES             Location/Qualifiers
     source          1. .11029
                     /organism="West Nile virus"
                     /mol_type="genomic RNA"
                     /isolate="WN NY 2000-grouse3282"
                     /specific_host="ruffed grouse"
                     /db_xref="taxon:11082"
                     /country="USA: New York"
     CDS             97. .10398
                     /note="contains capsid, pre-membrane, envelope, NS1, NS2a,
                     NS2b, NS3, NS4a, NS4b, and NS5"
                     /codon_start=1
                     /product="polyprotein precursor"
                     /protein_id="AAM81751.1"
                     /db_xref="GI:21929237"
                     /translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
                     RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
                     KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
                     AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
                     VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
                     VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
                     MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
                     NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQVG
                     ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
                     MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
                     NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
                     PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
                     INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
                     RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
                     ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
                     QMWEAVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
                     FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
                     KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
                     DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
                     TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
                     DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
                     GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
                     DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
                     LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
                     TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
                     ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMVCLAISAYTPWAILPSVV
                     GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
                     TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
                     VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
                     AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
                     LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
                     NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
                     TEIPDRAWNSGYEWITEYTGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
                     PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
                     SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
                     PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
                     TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
                     LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
                     VFFLLMQRKGIGKIGLGGAVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
                     QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
                     LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
                     GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
                     VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
                     TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
                     VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
                     RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
                     KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
                     CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
                     RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
                     PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
                     EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
                     EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
                     AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
                     YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
                     TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
                     RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
                     DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
                     QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
                     IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
                     DEKYVDYMSSLKRYEDTTLVEDTVL"
ORIGIN

  Query Match             100.0%;  Score 22;  DB 14;  Length 11029;
  Best Local Similarity   100.0%;  Pred. No. 1.5;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              ||||||||||||||||||||||
Db        195 AGCCCTCTTCAGTCCAATCAAG 174


RESULT 20
AF404756/c
LOCUS       AF404756               11029 bp ss-RNA     linear   VRL 23-JUL-2002
DEFINITION  West Nile virus isolate WN NY 2000-crow3356, complete genome.
ACCESSION   AF404756
VERSION     AF404756.1  GI:21929238
KEYWORDS    .
SOURCE      West Nile virus
  ORGANISM  West Nile virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE   1  (bases 1 to 11029)
  AUTHORS   Lanciotti,R.S., Ebel,G.D., Deubel,V., Kerst,A.J., Murri,S.,
            Meyer,R., Bowen,M., McKinney,N., Morrill,W.E., Crabtree,M.B.,
            Kramer,L.D. and Roehrig,J.T.
  TITLE     Complete genome sequences and phylogenetic analysis of West Nile
            virus strains isolated from the United States, Europe, and the
            Middle East
  JOURNAL   Virology 298 (1), 96-105 (2002)
  MEDLINE   22089180
   PUBMED   12093177
REFERENCE   2  (bases 1 to 11029)
  AUTHORS   Ebel,G.D., Kerst,A.J. and Lanciotti,R.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Division of Vector-Borne Infectious
            Diseases, Centers for Disease Control & Prevention, Rampart Road,
            Fort Collins, CO 80521, USA
FEATURES             Location/Qualifiers
     source          1. .11029
                     /organism="West Nile virus"
                     /mol_type="genomic RNA"
                     /isolate="WN NY 2000-crow3356"
                     /specific_host="crow"
                     /db_xref="taxon:11082"
                     /country="USA: New York"
     CDS             97. .10398
                     /note="contains capsid, pre-membrane, envelope, NS1, NS2a,
                     NS2b, NS3, NS4a, NS4b, and NS5"
                     /codon_start=1
                     /product="polyprotein precursor"
                     /protein_id="AAM81752.1"
                     /db_xref="GI:21929239"
                     /translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
                     RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
                     KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
                     AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
                     VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
                     VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
                     MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
                     NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQVG
                     ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
                     MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
```

NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
RLLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
QMWEAVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMVCLAISAYTPWAILPSVV
GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
TEIPDRAWNSGYEWITEYTGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
VFFLLMQRKGIGKIGLGGAVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
DEKYVDYMSSLKRYEDTTLVEDTVL"

ORIGIN

Query Match 100.0%; Score 22; DB 14; Length 11029;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCCCTCTTCAGTCCAATCAAG 22
||||||||||||||||||||||
Db 195 AGCCCTCTTCAGTCCAATCAAG 174

RESULT 21
AF481864/c
LOCUS AF481864 11029 bp RNA linear VRL 21-MAY-2002
DEFINITION West Nile virus strain IS-98 STD, complete genome.
ACCESSION AF481864
VERSION AF481864.1 GI:19387527
KEYWORDS .
SOURCE West Nile virus
ORGANISM West Nile virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Flavivirus; Japanese encephalitis virus group.
REFERENCE 1 (bases 1 to 11029)
AUTHORS Malkinson,M., Banet,C., Weisman,Y., Pokamunski,S., King,R.,
Drouet,M.T. and Deubel,V.
TITLE Introduction of West Nile virus in the Middle East by migrating
white storks
JOURNAL Emerging Infect. Dis. 8 (4), 392-397 (2002)
MEDLINE 21968420
PUBMED 11971773
REFERENCE 2 (bases 1 to 11029)
AUTHORS Deubel,V., Malkinson,M. and Banet,C.
TITLE Direct Submission
JOURNAL Submitted (08-FEB-2002) CERVI, Institut Pasteur, 21 Avenue Tony
Garnier, Lyon 69365, France
FEATURES Location/Qualifiers
source 1. .11029
/organism="West Nile virus"
/mol_type="genomic RNA"
/strain="IS-98 STD1"
/specific_host="sick stork"
/db_xref="taxon:11082"
/note="isolated in Israel in 1998"
CDS 97. .10398
/codon_start=1
/product="polyprotein precursor"
/protein_id="AAL87234.1"
/db_xref="GI:19387528"
/translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQVG
ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
ISRQELRCGNGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
QMWEAVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPRLRCLNLDVYRIL
LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
ESDAEITGSSERVDVRLDDGENFQLMNDPGAPWKIWMLRMVCLAISAYTPWAILPSVV
GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
TEIPDRAWNSGYEWITEYTGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMPDNINMPNGLIAQFYQ
PEREKVYTMEGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
VFFLLMQRKGIGKIGLGGAVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHSVSRGTAKLRWLVER
RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG

RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKINNRIERLRREYSSTWHHDENH
PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
EKVDTKAPEPPEGAKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
DEKYVDYMSSLKRYEDTTLVEDTVL"

ORIGIN

Query Match 100.0%; Score 22; DB 14; Length 11029;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCCCTCTTCAGTCCAATCAAG 22
||||||||||||||||||||||
Db 195 AGCCCTCTTCAGTCCAATCAAG 174

RESULT 22
AF533540/c
LOCUS AF533540 11029 bp RNA linear VRL 10-DEC-2002
DEFINITION West Nile virus polyprotein precursor, gene, complete cds.
ACCESSION AF533540
VERSION AF533540.1 GI:26284711
KEYWORDS .
SOURCE West Nile virus (WNV)
ORGANISM West Nile virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Flavivirus; Japanese encephalitis virus group.
REFERENCE 1 (bases 1 to 11029)
AUTHORS Huang,C., Slater,B., Rudd,R., Parchuri,N., Hull,R., Dupuis,M. and Hindenburg,A.
TITLE First Isolation of West Nile virus from a Patient with Encephalitis in the United States
JOURNAL Emerging Infect. Dis. 8 (12), 1367-1371 (2002)
REFERENCE 2 (bases 1 to 11029)
AUTHORS Huang,C., Slater,B., Rudd,R., Parchuri,N., Hull,R., Dupuis,M. and Hindenburg,A.
TITLE Direct Submission
JOURNAL Submitted (30-JUL-2002) Wadsworth Center, New York State Department of Health, Box 509, Albany, NY 12201-0509, USA
FEATURES Location/Qualifiers
source 1. .11029
/organism="West Nile virus"
/mol_type="genomic RNA"
/isolation_source="human cerebrospinal fluid, New York, 2001"
/db_xref="taxon:11082"
/country="USA"
CDS 97. .10398
/codon_start=1
/product="polyprotein precursor"
/protein_id="AAN85090.1"
/db_xref="GI:26284712"
/translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQVG
ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID

ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
QMWEAVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
GAAFAESNSGGDVVHLALMATFKIQPVFVVASFLKARWTNQENILLMLAAVFFQMAYH
DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMVCLAISAYTPWAILPSVV
GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
AIVQGERMDEPIPAGFEPEMLRKKKITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
TEIPDRAWNSGYEWITEYTGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
VFFLLMQRKGIGKIGLGGAVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
DEKYVDYMSSLKRYEDTTLVEDTVL"
mat_peptide 97. .465
/product="nucleocapsid protein"
mat_peptide 466. .741
/product="pre-membrane protein M"
mat_peptide 742. .966
/product="membrane protein M"
mat_peptide 967. .2469
/product="envelope glycoprotein E"
mat_peptide 2470. .3525
/product="non-structural protein 1"
mat_peptide 3526. .4218
/product="non-structural protein 2A"
mat_peptide 4219. .4611
/product="non-structural protein 2B"
mat_peptide 4612. .6468
/product="non-structural protein 3"
mat_peptide 6469. .6915
/product="non-structural protein 4A"
mat_peptide 6916. .7680
/product="non-structural protein 4B"
mat_peptide 7681. .10395
/product="non-structural protein 5"

ORIGIN

Query Match 100.0%; Score 22; DB 14; Length 11029;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              ||||||||||||||||||||||
Db        195 AGCCCTCTTCAGTCCAATCAAG 174


RESULT 23
AY289214/c
LOCUS       AY289214               11029 bp    RNA     linear   VRL 01-JUN-2004
DEFINITION  West Nile virus strain TVP 8533 complete genome.
ACCESSION   AY289214
VERSION     AY289214.1  GI:33948906
KEYWORDS    .
SOURCE      West Nile virus (WNV)
  ORGANISM  West Nile virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE   1  (bases 1 to 11029)
  AUTHORS   Granwehr,B.P., Li,L., Davis,C.T., Beasley,D.W.C. and Barrett,A.D.T.
  TITLE     Phylogenetic Analysis of a Human Isolate of West Nile Virus from
            the Gulf Coast of Texas, 2002
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 11029)
  AUTHORS   Granwehr,B.P. and Barrett,A.D.T.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-2003) Internal Medicine-Infectious Diseases,
            University of Texas Medical Branch, 301 University Blvd., Rte.
            0435, Galveston, TX 77555/0435, USA
FEATURES             Location/Qualifiers
     source          1. .11029
                     /organism="West Nile virus"
                     /mol_type="genomic RNA"
                     /strain="TVP 8533"
                     /db_xref="taxon:11082"
                     /note="Beaumont Strain"
     5'UTR           1. .96
     CDS             97. .10398
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="AAQ55854.1"
                     /db_xref="GI:33948907"
                     /translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
                     RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
                     KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
                     AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
                     VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
                     VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
                     MMNMEAANLAEVRSYCYLATVSDLSTKAACPAMGEAHNDKRADPAFVCRQGVVDRGWG
                     NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQVG
                     ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
                     MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
                     NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
                     PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
                     INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
                     RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
                     ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
                     QMWEAVKDELNTLLKENGVDLSVVVEKQGGMYKSAPKRLTATTEKLEIGWKAWGKSIL
                     FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
                     KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
                     DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
                     TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
                     DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
                     GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
                     DARQILLWEIPDVLNSLAIAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
                     LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
                     TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
                     ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMVCLAISAYTPWAILPSVV
                     GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
                     TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
                     VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
                     AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
                     LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
                     NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
                     TEIPDRAWNSGYEWITEYTGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
                     PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
                     SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
                     PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
                     TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
                     LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
                     VFFLLMQRKGIGKIGLGGAVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
                     QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
                     LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
                     GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
                     VVDGIVATDVPELERTTPIMQKKIGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
                     TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
                     VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
                     RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
                     KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
                     CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
                     RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
                     PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
                     EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
                     EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
                     AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGIRPGGKI
                     YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
                     TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
                     RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
                     DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
                     QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
                     IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
                     DEKYVDYMSSLKRYEDTTLVEDTVL"
     mat_peptide     97. .465
                     /product="nucleocapsid protein"
                     /note="C"
     mat_peptide     466. .741
                     /product="pre-membrane protein M"
                     /note="prM"
     mat_peptide     742. .966
                     /product="membrane protein M"
                     /note="M"
     mat_peptide     967. .2469
                     /product="envelope glycoprotein E"
                     /note="E"
     mat_peptide     2470. .3525
                     /product="non-structural protein 1"
                     /note="NS1"
     mat_peptide     3526. .4218
                     /product="non-structural protein 2A"
                     /note="NS2A"
     mat_peptide     4219. .4611
                     /product="non-structural protein 2B"
                     /note="NS2B"
     mat_peptide     4612. .6468
                     /product="non-structural protein 3"
                     /note="NS3"
     mat_peptide     6469. .6915
                     /product="non-structural protein 4A"
                     /note="NS4A"
     mat_peptide     6916. .7680
                     /product="non-structural protein 4B"
                     /note="NS4B"
     mat_peptide     7681. .10395
                     /product="non-structural protein 5"
                     /note="NS5"
     3'UTR           10399. .11029
ORIGIN

  Query Match           100.0%;  Score 22;  DB 14;  Length 11029;
  Best Local Similarity  100.0%;  Pred. No. 1.5;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              ||||||||||||||||||||||
Db        195 AGCCCTCTTCAGTCCAATCAAG 174
```

```
RESULT 24
AY688948/c
LOCUS       AY688948               11057 bp    RNA     linear   VRL 15-AUG-2004
DEFINITION  West Nile virus strain Sarafend, complete genome.
ACCESSION   AY688948
VERSION     AY688948.1  GI:51095221
KEYWORDS    .
SOURCE      West Nile virus (WNV)
  ORGANISM  West Nile virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE   1  (bases 1 to 11057)
  AUTHORS   Li,J., Bhuvanakantham,R. and Ng,M.-L.
  TITLE     Construction and characterization of an infectious West Nile
            (Sarafend) clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 11057)
  AUTHORS   Li,J., Bhuvanakantham,R. and Ng,M.-L.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUL-2004) Microbiology, National University of
            Singapore, 5 Science Drive 2, Singapore 117597, Singapore
FEATURES             Location/Qualifiers
     source          1. .11057
                     /organism="West Nile virus"
                     /virion
                     /mol_type="genomic RNA"
                     /strain="Sarafend"
                     /db_xref="taxon:11082"
                     /country="Israel"
                     /note="lineage II"
     CDS             97. .10401
                     /codon_start=1
                     /product="putative polyprotein precursor"
                     /protein_id="AAT95390.1"
                     /db_xref="GI:51095222"
                     /translation="MSKKPGGPGKNRAVNMLKRGMPRGLSLIGLKRAMLSLIDGKGPI
                     RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSTKQ
                     KKRGGTAGFTILLGLIACAGAVTLSNFQGKVMMTVNATDVTDAITIPTAAGKNLCIVR
                     AMDVGYLCEDTITYECPVLAAGNDPEDIDCWCTKSSVYVRYGRCTKTRHSRRSRRSLT
                     VQTHGESTLANKKGAWLDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
                     VFAILLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
                     MMNMEAANLADVRSYCYLASVSDLSTKAACPTMGEAHNEKRADPAFVCKQGVVDRGWG
                     NGCGLFGKGSIDTCAKFACTTKATGWIIQKENIKYEVAIFVHGPTTVESHGNYSTQIG
                     ATQAGRFSITPSAPSYTLKLGEYGEVTVDCEPRSGIDTSAYYVMSVGAKSFLVHREWF
                     MDLNLPWSSAGSTTWRNRETLMEFEEPHATKQSVVALGSQEGALHQALAGAIPVEFSS
                     NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFARTPADTGHGTVVLELQYTGKDG
                     PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANSKVLIELEPPFGDSYIVVGRGEQQ
                     INHHWHKSGSSIGKAFTTTLRGAQRLAALGDTAWDFGSVGGVFTSVGKAIHQVFGGAF
                     RSLFGGMSWITQGLLGALLLWMGINARDRSIAMTFLAVGGVLLFLSVNVHADTGCAID
                     VGRQELRCGSGVFIHNDVEAWMDRYKFYPETPQGLAKIIQKAHAEGVCGLRSVSRLEH
                     QMWEAIKDELNTLLKENGVDLSVVVEKQNGMHKAAPRRLAATTEKLEMGWKAWGKSII
                     FAPELANNTFVIDGPETEECPTANRAWNSMEVEDFGFGLTSTRMFLKIRETNTTECDS
                     KIIGTAVKNNMAVHSDLSYWIESGLNDTWKLERAVLGEVKSCTWPETHTLWGDGVLES
                     DLIIPITLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTISDSCGHRGPA
                     ARTTTESGKLITDWCCRSCTLPPLRFKTENGCWYGMEIRPLRHDEKTLVQSRVNAYNA
                     DMIDPFQLGLLIVFLATQEVLRKRWTAKISMPAIMLALLVLVFGGITYTDVLRYVILV
                     GAAFAEANSGGDVVHLALMATFKIQPVFLVASFLKARWTNQESILLMLAAAFFQMAYY
                     DAKNVLAWEMPDVLNSLSVAWMILRAISFTNTSNVVVPLLALLTPGLKCLNLDVYRIL
                     LLMVGVGSLIKEKRSSAAKKKGACLICLALASTGVFNPMILAAGLMACDPNRKRGWPA
                     TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFVAFVISGKSTDMWIERTADITW
                     ESDAEITGSSERVDVKLDDDGNFQLMNDPGAPWKIWMLRMACLAISAYTPWAILPSVI
                     GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
                     TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGHDEVQMIVVEPGKN
                     VKNVQTKPGVFKTPEGEIGAVTLDYPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
                     AIVQGERMEEPAPAGFEPEMLRKKQITVLDLHPGAGKTRKILPQIIKEAINRRLRTAV
                     LAPTRVVAAEMSEALRGLPIRYQTSAVHREHSGNEIVDVMCHATLTHRLMSPHRVPNY
                     NLFIMDEAHFTDPASIAARGYIATKVELGEAAAIFMTATPPGTSDPFPESNAPISDMQ
                     TEIPDRAWNTGYEWITEYVGKTVWFVPSVKMGNEIALCLQRAGKKVIQLNRKSYETEY
                     PKCKNDDWDFVITTDISEMGANFKANRVIDSRKSVKPTIIEEGDGRVILGEPSAITAA
                     SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLVAQLYQ
                     PEREKVYTMDGEYRLRGEERKNFLEFLRTADLPVWLAYKVAAAGISYHDRKWCFDGPR
                     TNTILEDNNEVEVITKLGERKILRPRWVDARVYSDHQALKSFKDFASGKRSQIGLVEV
                     LGRMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALITLLSVMSLG
                     VFFLLMQRKGIGKIGLGGVILGAATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
                     QRSQTDNQLAVFLICVLTLVSAVAANEMGWLDKTKSDIGSLLGHKPEARETTLGVESF
                     LLDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVD
                     VGVSALLLAAGCWGQVTLTVTVTAAALLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKN
                     AVVDGIVATDVPELERTTPVMQKKVGQIMLILVSMAAVVVNPSVRTVREAGILTTAAA
                     VTLWENGASSVWNATTAIGLCHIMRGGWLSCLSIAWTLIKSMEKPVLKRGGAKGRTLG
                     EVWKERLNHMTKEEFTRYRKEAITEVDRSAAKHARREGNITGGHPVSRGTAKLRWLVE
                     RRFLEPVGKVVDLGCGRGGWCYYMATQKRVQEVKGYTKGGPGHEEPQLVQSYGWNIVT
                     MKSGVDVFYRPSEASDTLLCDIGESSSSAEVEEHRTVRVLEMVEDWLHRGPKEFCIKV
                     LCPYMPKVIEKMETLQRRYGGGLVRNPLSRNSTHEMYWVSHASGNIVHSINMTSQVLL
                     GRMEKKTWKGPQYEEDVNLGSGTRAVGRPLLNSDTSKIKNRIERLKKEYNSTWHQDVN
                     HPYRTWNYHGSYEVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVF
                     KEKVDTKAPEPPEGVKYVLNETTNWLWAFLARDKKPRMCSREEFIGKVNSNAALGAMF
                     EEQNQWKNAREAVEDPKFWEMVDEEREAHLRGECNTCIYNMMGKREKKPGEFGKAKGS
                     RAIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILKEVGTKPGGK
                     TYADDTAGWDTRITKADLENEAKVLELLDGEHRRLARSIIELTYRHKVVKVMRPAADG
                     KTVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLGKGKGPK
                     VRTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGW
                     YDWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELIGRARISPGAGWNVRDTACLAKSY
                     AQMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAKGEWMTTEDMLAVWNRV
                     WIEENEWMEDKTPVERWSDVPYSGKREDIWCGSLIGTRTRATWAENIHVAINQVRSVI
                     GEEKYVDYMGSLRRYEDTTVVEDTVL"
     mat_peptide     97. .465
                     /product="putative anchored core protein C"
     mat_peptide     466. .966
                     /product="putative preM protein"
     mat_peptide     742. .966
                     /product="putative matrix protein M"
     mat_peptide     967. .2469
                     /product="putative envelope protein"
     mat_peptide     2470. .3525
                     /product="putative non-structural protein NS1"
     mat_peptide     3526. .4218
                     /product="putative non-structural protein NS2a"
     mat_peptide     4219. .4611
                     /product="putative non-structural protein NS2b"
     mat_peptide     4612. .6468
                     /product="putative non-structural protein NS3"
     mat_peptide     6469. .6846
                     /product="putative non-structural protien NS4a"
     mat_peptide     6847. .6915
                     /product="putative 2k protein"
     mat_peptide     6916. .7683
                     /product="putative non-structural protein NS4b"
     mat_peptide     7684. .10398
                     /product="putative non-structural protein NS5"
ORIGIN

  Query Match           100.0%;  Score 22;  DB 14;  Length 11057;
  Best Local Similarity  100.0%;  Pred. No. 1.5;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCCCTCTTCAGTCCAATCAAG 22
              ||||||||||||||||||||||
Db        195 AGCCCTCTTCAGTCCAATCAAG 174


RESULT 25
AF375043/c
LOCUS       AF375043               1648 bp    mRNA    linear   VRL 14-MAR-2002
DEFINITION  West Nile virus isolate WN_0233 polyprotein mRNA, partial cds.
ACCESSION   AF375043
VERSION     AF375043.1  GI:19421849
KEYWORDS    .
SOURCE      West Nile virus
  ORGANISM  West Nile virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE   1  (bases 1 to 1648)
  AUTHORS   Hindiyeh,M., Shulman,L.M., Mendelson,E., Weiss,L., Grossman,Z. and
            Bin,H.
  TITLE     Isolation and characterization of West Nile virus from the blood of
            viremic patients during the 2000 outbreak in Israel
```

```
  JOURNAL   Emerging Infect. Dis. 7 (4), 748-750 (2001)
  MEDLINE   21469825
   PUBMED   11585544
REFERENCE   2  (bases 1 to 1648)
  AUTHORS   Hindiyeh,M., Shulman,L.M., Mendelson,E., Grossman,Z., Weiss,L. and
            Bin,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-APR-2001) Central Virology Laboratory, Ministry of
            Health, Public Health Laboratories, Sheba Medical Center, Tel
            Hashomer 52621, Israel
FEATURES             Location/Qualifiers
     source          1. .1648
                     /organism="West Nile virus"
                     /mol_type="mRNA"
                     /isolate="WN_0233"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:11082"
     CDS             <1. .>1648
                     /codon_start=3
                     /product="polyprotein"
                     /protein_id="AAL87746.1"
                     /db_xref="GI:19421850"
                     /translation="SLIGLKRAMLSLIDGKGPIRFVLALLAFFRFTAIAPTRAVLDRW
                     RGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQKKRGGKIGIAVMIGLIASVGAVTLS
                     NFQGKVMMTVNATDVTDVITIPTAAGKNLCIVRAMDVGYMCDDTITYECPVLSAGNDP
                     EDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLTVQTHGESTLANKKGAWMDSTKATRY
                     LVKTESWILRNPGYALVAAVIGWMLGSNTMQRVVFVVLLLLVAPAYSFNCLGMSNRDF
                     LEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLAEVRSYCYLATVSDLS
                     TKAACPTMGEAHNDKRADPAFVCKQGVVDRGWGNGCGLFGKGSIDTCAKFACSTKATG
                     RTILKENIKYEVAIFVHGPTTVESHGNYSTQIGATQAGRFSITPAAPSHTLKLGEYGE
                     VTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWFMDLNLPWSSAGSTVWRNRETLMEFE
                     EPHATKQSVIALGSQEGALHQALAGAIPVEFSSNTVKLTSG"
ORIGIN

  Query Match             95.5%;  Score 21;  DB 14;  Length 1648;
  Best Local Similarity   100.0%;  Pred. No. 5.1;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GCCCTCTTCAGTCCAATCAAG 22
              |||||||||||||||||||||
Db         25 GCCCTCTTCAGTCCAATCAAG 5


Search completed: September  6, 2005, 20:29:45
Job time : 775.688 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 6, 2005, 16:01:23 ; Search time 735.656 Seconds
(without alignments)
1383.200 Million cell updates/sec

Title: US-10-729-421-34
Perfect score: 21
Sequence: 1 ccgggctgtcaatatgctaaa 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database :
GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	21	100.0	240	14	WNF42SAA	M32560 West Nile v
2	21	100.0	2440	14	AF194117	AF194117 West Nile
3	21	100.0	10664	14	KUNCG	D00246 Kunjin viru
4	21	100.0	10842	14	AY278442	AY278442 West Nile
5	21	100.0	10845	14	AY277252	AY277252 West Nile
6	21	100.0	10945	14	AF202541	AF202541 West Nile
7	21	100.0	10962	14	WNFCG	M12294 West Nile v
8	21	100.0	10972	14	AF317203	AF317203 West Nile
9	21	100.0	10975	14	AF206518	AF206518 West Nile
10	21	100.0	10984	14	AY262283	AY262283 West Nile
11	21	100.0	10989	14	AY268132	AY268132 West Nile
12	21	100.0	10989	14	AY268133	AY268133 West Nile
13	21	100.0	11022	14	AY274504	AY274504 Kunjin vi
14	21	100.0	11022	14	AY274505	AY274505 Kunjin vi
15	21	100.0	11028	14	AY490240	AY490240 West Nile
16	21	100.0	11029	6	AX576542	AX576542 Sequence
17	21	100.0	11029	6	AX577796	AX577796 Sequence
18	21	100.0	11029	14	AB185914	AB185914 West Nile
19	21	100.0	11029	14	AB185915	AB185915 West Nile
20	21	100.0	11029	14	AB185916	AB185916 West Nile
21	21	100.0	11029	14	AB185917	AB185917 West Nile
22	21	100.0	11029	14	AF196835	AF196835 West Nile
23	21	100.0	11029	14	AF260967	AF260967 West Nile
24	21	100.0	11029	14	AF260968	AF260968 West Nile
25	21	100.0	11029	14	AF260969	AF260969 West Nile
26	21	100.0	11029	14	AF404753	AF404753 West Nile
27	21	100.0	11029	14	AF404754	AF404754 West Nile
28	21	100.0	11029	14	AF404755	AF404755 West Nile
29	21	100.0	11029	14	AF404756	AF404756 West Nile
30	21	100.0	11029	14	AF404757	AF404757 West Nile
31	21	100.0	11029	14	AF481864	AF481864 West Nile
32	21	100.0	11029	14	AF533540	AF533540 West Nile
33	21	100.0	11029	14	AY289214	AY289214 West Nile
34	21	100.0	11057	14	AY688948	AY688948 West Nile
35	20	95.2	10998	14	AY278441	AY278441 West Nile
36	19.4	92.4	4673	14	SLOCME	M16614 St. Louis e
37	19.4	92.4	10741	14	AY277251	AY277251 West Nile
38	18	85.7	112979	5	BX088699	BX088699 Zebrafish
39	17.8	84.8	2379	14	S75726	S75726 Japanese en
40	17.8	84.8	2435	14	JEU02367	U02367 Japanese en
41	17.8	84.8	2435	14	JEU21054	U21054 Japanese en
42	17.8	84.8	2436	14	JEU03689	U03689 Japanese en
43	17.8	84.8	2436	14	JEU03691	U03691 Japanese en
44	17.8	84.8	2436	14	JEU03692	U03692 Japanese en
45	17.8	84.8	2436	14	JEU03693	U03693 Japanese en
46	17.8	84.8	2436	14	JEU03694	U03694 Japanese en
47	17.8	84.8	2436	14	JEU03695	U03695 Japanese en
48	17.8	84.8	2477	14	JEU04521	U04521 Japanese en
49	17.8	84.8	2477	14	JEU04522	U04522 Japanese en
50	17.8	84.8	3000	14	JEV5END	M15337 Japanese en
51	17.8	84.8	4406	14	AY027863	AY027863 Japanese
52	17.8	84.8	4512	6	AR003118	AR003118 Sequence
53	17.8	84.8	4512	6	AR003211	AR003211 Sequence
54	17.8	84.8	4512	14	JEVCPRMEN	M73710 Japanese en
55	17.8	84.8	5436	14	FLMVEV5	X03467 Murray Vall
56	17.8	84.8	10951	14	JEVLINGCG	L78128 Japanese en
57	17.8	84.8	10964	14	AF217620	AF217620 Japanese
58	17.8	84.8	10968	14	AY585242	AY585242 Japanese
59	17.8	84.8	10968	14	AY585243	AY585243 Japanese
60	17.8	84.8	10969	14	JEU15763	U15763 Japanese en
61	17.8	84.8	10969	14	JEVCMNAA	M55506 Japanese en
62	17.8	84.8	10970	14	AF254452	AF254452 Japanese
63	17.8	84.8	10970	14	AF254453	AF254453 Japanese
64	17.8	84.8	10970	14	AY303791	AY303791 Japanese
65	17.8	84.8	10970	14	AY303792	AY303792 Japanese
66	17.8	84.8	10970	14	AY303793	AY303793 Japanese
67	17.8	84.8	10970	14	AY303794	AY303794 Japanese
68	17.8	84.8	10970	14	AY303795	AY303795 Japanese
69	17.8	84.8	10970	14	AY303796	AY303796 Japanese
70	17.8	84.8	10970	14	AY303797	AY303797 Japanese
71	17.8	84.8	10970	14	AY303798	AY303798 Japanese
72	17.8	84.8	10976	14	AF014160	AF014160 Japanese
73	17.8	84.8	10976	14	AF014161	AF014161 Japanese
74	17.8	84.8	10976	14	AF075723	AF075723 Japanese
75	17.8	84.8	10976	14	AF098735	AF098735 Japanese
76	17.8	84.8	10976	14	AF098736	AF098736 Japanese
77	17.8	84.8	10976	14	AF098737	AF098737 Japanese
78	17.8	84.8	10976	14	AF221499	AF221499 Japanese
79	17.8	84.8	10976	14	AF221500	AF221500 Japanese
80	17.8	84.8	10976	14	AF315119	AF315119 Japanese
81	17.8	84.8	10976	14	AF416457	AF416457 Japanese
82	17.8	84.8	10976	14	JEU14163	U14163 Japanese en
83	17.8	84.8	10976	14	JEU47032	U47032 Japanese en
84	17.8	84.8	10976	14	JEVBEICG	L48961 Japanese en
85	17.8	84.8	10976	14	JEVCG	M18370 Japanese en
86	17.8	84.8	10976	14	JEVSAA	D90195 Japanese en
87	17.8	84.8	10976	14	JEVSAV	D90194 Japanese en
88	17.8	84.8	10977	14	AF069076	AF069076 Japanese
89	17.8	84.8	10977	14	AF080251	AF080251 Japanese
90	17.8	84.8	10977	14	AF486638	AF486638 Japanese
91	17.8	84.8	11014	14	AF161266	AF161266 Murray Va
c 92	17.8	84.8	155933	2	AC141389	AC141389 Rattus no

c	93	17.8	84.8	170073	9	AC060234	AC060234 Homo sapi
c	94	17.4	82.9	15326	8	SPAC140	AL163191 S.pombe c
c	95	17.4	82.9	150121	10	AC114925	AC114925 Mus muscu
	96	17.4	82.9	197486	2	AC145692	AC145692 Mus muscu
c	97	17.4	82.9	199087	2	CR848021	CR848021 Danio rer
	98	17.4	82.9	200081	2	AC107693	AC107693 Mus muscu
c	99	17.4	82.9	200890	5	BX465192	BX465192 Zebrafish
	100	17.4	82.9	204314	5	BX465854	BX465854 Zebrafish

ALIGNMENTS

```
RESULT 1
WNF42SAA
LOCUS       WNF42SAA                 240 bp ss-RNA     linear   VRL 06-JUL-1995
DEFINITION  West Nile virus (WN) 5' terminal region of genome.
ACCESSION   M32560
VERSION     M32560.1  GI:336165
KEYWORDS    .
SOURCE      West Nile virus
  ORGANISM  West Nile virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE   1  (bases 1 to 240)
  AUTHORS   Castle,E. and Wengler,G.
  TITLE     Nucleotide sequence of the 5'-terminal untranslated part of the
            genome of the flavivirus West Nile virus
  JOURNAL   Arch. Virol. 922, 309-313 (1987)
  MEDLINE   87127557
COMMENT     Original source text: West Nile virus cDNA to genomic RNA.
FEATURES             Location/Qualifiers
     source          1. .240
                     /organism="West Nile virus"
                     /mol_type="genomic RNA"
                     /db_xref="taxon:11082"
     CDS             97. .>240
                     /note="ORF A; putative"
                     /codon_start=1
                     /product="unknown protein"
                     /protein_id="AAA69639.1"
                     /db_xref="GI:893350"
                     /translation="MSKKPGGPGKNRAVNMLKRGMPRGLSLIGLKRAMLSLIDGKGPI
                     RFVL"
     CDS             142. .>240
                     /note="ORF B; putative"
                     /codon_start=1
                     /product="unknown protein"
                     /protein_id="AAA69640.1"
                     /db_xref="GI:893351"
                     /translation="MLKRGMPRGLSLIGLKRAMLSLIDGKGPIRFVL"
ORIGIN

  Query Match             100.0%;  Score 21;  DB 14;  Length 240;
  Best Local Similarity   100.0%;  Pred. No. 1.1;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCTAAA 21
              |||||||||||||||||||||
Db        129 CCGGGCTGTCAATATGCTAAA 149


RESULT 2
AF194117
LOCUS       AF194117                 2440 bp    RNA     linear   VRL 19-JAN-2000
DEFINITION  West Nile virus structural protein precursor, gene, partial cds.
ACCESSION   AF194117
VERSION     AF194117.1  GI:6715269
KEYWORDS    .
SOURCE      West Nile virus
  ORGANISM  West Nile virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE   1  (sites)
  AUTHORS   Lanciotti,R.S., Roehrig,J.T., Deubel,V., Smith,J., Parker,M.,
            Steele,K., Crise,B., Volpe,K.E., Crabtree,M.B., Scherret,J.H.,
            Hall,R.A., MacKenzie,J.S., Cropp,C.B., Panigrahy,B., Ostlund,E.,
            Schmitt,B., Malkinson,M., Banet,C., Weissman,J., Komar,N.,
            Savage,H.M., Stone,W., McNamara,T. and Gubler,D.J.
  TITLE     Origin of the West Nile virus responsible for an outbreak of
            encephalitis in the northeastern United States
  JOURNAL   Science 286 (5448), 2333-2337 (1999)
  MEDLINE   20070288
   PUBMED   10600742
REFERENCE   2  (bases 1 to 2440)
  AUTHORS   Parker,M.D., Crise,B.J., Clayton,J.M. and Smith,J.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1999) Virology Division, U.S. Army Medical
            Research Institute of Infectious Diseases, Bldg. 1425 Fort Detrick,
            Frederick, Maryland 21702, USA
FEATURES             Location/Qualifiers
     source          1. .2440
                     /organism="West Nile virus"
                     /virion
                     /mol_type="genomic RNA"
                     /db_xref="taxon:11082"
     CDS             1. .>2440
                     /codon_start=1
                     /product="structural protein precursor"
                     /protein_id="AAF26360.1"
                     /db_xref="GI:6715270"
                     /translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
                     RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
                     KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
                     AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
                     VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
                     VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
                     MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
                     NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQVG
                     ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
                     MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
                     NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
                     PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
                     INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
                     RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
                     ISRQELRCGSGVFIH"
ORIGIN

  Query Match             100.0%;  Score 21;  DB 14;  Length 2440;
  Best Local Similarity   100.0%;  Pred. No. 1.4;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCTAAA 21
              |||||||||||||||||||||
Db         33 CCGGGCTGTCAATATGCTAAA 53


RESULT 3
KUNCG
LOCUS       KUNCG                    10664 bp    RNA     linear   VRL 17-FEB-1998
DEFINITION  Kunjin virus gene for polyprotein (C, prM, E, NS1, NS2A, NS2B, NS3,
            NS4A, NS4B, NS5), complete cds.
ACCESSION   D00246
VERSION     D00246.1  GI:221966
KEYWORDS    M (membrane protein); prM (precursor of M); NS5; NS4B; NS4A; NS3;
            NS2B; NS2A; NS1; E (envelope protein); C (core protein);
            polyprotein.
SOURCE      Kunjin virus
  ORGANISM  Kunjin virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE   1  (bases 1 to 10664)
  AUTHORS   Coia,G., Parker,M.D., Speight,G., Byrne,M.E. and Westaway,E.G.
  TITLE     Nucleotide and complete amino acid sequences of Kunjin virus:
            definitive gene order and characteristics of the virus-specified
            proteins
```

```
  JOURNAL   J. Gen. Virol. 69 (Pt 1), 1-21 (1988)
  MEDLINE   88089524
   PUBMED   2826659
COMMENT     A kunjin (KUN) virus cDNA sequence of 10664 nucleotides which
            encoded a single open reading frame for 3433 amino acids was
            obtained and compared with the complete amino acid sequences of
            yellow fever and West Nile viruses.  Partial N-terminal amino acid
            analyses of KUN virus-specified proteins identified the polyprotein
            cleavage sites and the definitive gene order.  Three stop codons in
            the correct reading frame occur within the first 25 nucleotides
            beyond the 3' end of the coding sequence.
FEATURES             Location/Qualifiers
     source          1. .10664
                     /organism="Kunjin virus"
                     /mol_type="genomic RNA"
                     /strain="MRM61C"
                     /db_xref="taxon:11077"
                     /clone="pKV479, pKV681, pKV87, pKV761, pKV62, pKV71,
                     pKV621, pKV75, pKV90, PKV66"
     CDS             76. .10377
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="BAA00176.1"
                     /db_xref="GI:221967"
                     /translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLTGLKRAMLSLIDGRGPT
                     RFVLALLAFFRFTAIAPTRAVLDRWRSVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
                     KKRGGKTGIAFMIGLIAGVGAVTLSNFQGKVMMTVNATDVTDIITIPPAAGKNLCIVR
                     AMDVGHMCDDTITYECPVLSAGNDPEDIDCWCTKLAVYVRYGRCTKTRHSRRSRRSLT
                     VQTHGESTLSNKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
                     VFAVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
                     MMNMEAANLAEVRSYCYLATVSELSTKAACPTMGEAHNDKRADPSFVCKQGVVDRGWG
                     NGCGLFGKGSIDTCAKFACSTKATGRTILKENIKYEVAIFVHGPTTVESHGNYFTQTG
                     AAQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTSAYYVMTVGTKTFLVHREWF
                     MDLNLPWSSAESNVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
                     NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFRFLGTPADTGHGTVVLELQYTGTDG
                     PCKIPISSVASLNDLTPVGRLVTVNPFVSVSTANAKVLIELEPPFGDSYIVVGRGEQQ
                     INHHWHKSGSSIGKAFTATLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
                     RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
                     ISRQELRCGSGVFIHNDVEAWIDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
                     QMWEAVKDELNTLLKENGVDLSIVVEKQEGMYKSAPRRLTATTEKLEIGWKAWGKSIL
                     FAPELANNTFVIDGPETKECPTQNRAWNNLEVEDFGFGLTSTRMFLRVRESNTTECDS
                     KIIGTAVKNNLAIHSDLSYWIESRFNDTWKLERAVLGEVKSCTWPETHTLWGDGVLES
                     DLIIPITLAGPRSNHNRRPGYKTQSQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
                     TRTTTESGKLITDWCCRSCTLPPLRYQTDNGCWYGMEIRPQRHDEKTLVQSQVNAYNA
                     DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
                     GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAAFFQMAYY
                     DARQILLWEMPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
                     LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGFFNPMILAAGLVACDPNRKRGWPA
                     TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
                     EGDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMACLAISAYTPWAILPSVV
                     GFWITLQYTKRGGVLWDTPSPKEYKRGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
                     TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
                     VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
                     AIVQGERMDEPVPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
                     LAPTRVVAAEMAEALRGLPIRYQTSAVAREHNGNEIVDVMCHATLTHRLMSPHRVPNY
                     NLFVMDEAHFTDPASIAARGYISTRVELGEAAAIFMTATPPGTSDPFPESNAPISDLQ
                     TEIPDRAWNSGYEWITEYIGKTVWFVPSVKMGNEIALCLQRAGKKVIQLNRKSYETEY
                     PKCKNDDWDFVVTTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
                     SAAQRRGRTGRNPSQAGDEYCYGGHTNEDDSNCAHWTEARIMLDNINMPNGLIAQFYQ
                     PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
                     TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKSFKDFASGKRSQIGFIEV
                     LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
                     VFFLLMQRKGIGKIGLGGVVLGAATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
                     QRSQTDNQLAVFLICVLTLVGAVAANEMGWLDKTKSDISGLFGQRIETKENFSIGEFL
                     LDLRPATAWSLYAVTTAVLTPLLKHLITSDYITTSLTSINVQASALFTLARGFPFVDV
                     GVSALLLAAGCWGQVTLTVTVTSATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
                     VVDGIVATDVPELERTTPIMQKKVGQVMLILVSLAALVVNPSVKTVREAGILITAAAV
                     TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLVKNMEKPGLKRGGAKGRTLGE
                     VWKERLNQMTKEEFIRYRKEAITEVDRSAAKHARKERNITGGHPVSRGTAKLRWLVER
                     RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
                     KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTLRVLEMVEDWLHRGPKEFCVKVL
                     CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
                     RMEKKTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
                     PYRTWNYHGSYEVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
                     EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
                     EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
                     AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGRI
                     YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
                     TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
                     RTWLSENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
                     DWQQVPFCSNHFTELIMKDGRTLVTPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
                     QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
                     IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRSIIG
                     DEKYVDYMSSLKRYEDTTLVEDTVL"
     mat_peptide     79. .414
                     /product="C (core protein)"
     mat_peptide     415. .945
                     /product="prM (precursor of M)"
     mat_peptide     721. .945
                     /product="M (membrane protein)"
     mat_peptide     946. .2448
                     /product="E (envelope protein)"
     mat_peptide     2449. .3504
                     /product="NS1"
     mat_peptide     3505. .4197
                     /product="NS2A"
     mat_peptide     4198. .4590
                     /product="NS2B"
     mat_peptide     4591. .6447
                     /product="NS3"
     mat_peptide     6448. .6894
                     /product="NS4A"
     mat_peptide     6895. .7659
                     /product="NS4B"
     mat_peptide     7660. .10374
                     /product="NS5"
ORIGIN

  Query Match           100.0%;  Score 21;  DB 14;  Length 10664;
  Best Local Similarity  100.0%;  Pred. No. 1.6;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCTAAA 21
              |||||||||||||||||||||
Db        108 CCGGGCTGTCAATATGCTAAA 128


RESULT 4
AY278442
LOCUS       AY278442               10842 bp    RNA     linear   VRL 03-MAY-2003
DEFINITION  West Nile virus isolate LEIV-Vlg00-27924, complete genome.
ACCESSION   AY278442
VERSION     AY278442.1  GI:30349731
KEYWORDS    .
SOURCE      West Nile virus (WNV)
  ORGANISM  West Nile virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE   1  (bases 1 to 10842)
  AUTHORS   Sadykova,G.K., Prilipov,A.G., Kinney,R.M., Samokhvalov,E.I.,
            Savage,H.M., Alkhovsky,S.V., Tsychia,R., Gromashevsky,V.L.,
            Usachev,E.V., Mokhonov,V.V., Voronina,A.G., Butenko,A.M.,
            Larichev,V.F., Gubler,D.J. and Lvov,D.K.
  TITLE     Analysis of a new variants of West Nile virus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 10842)
  AUTHORS   Sadykova,G.K., Prilipov,A.G., Kinney,R.M., Samokhvalov,E.I.,
            Savage,H.M., Alkhovsky,S.V., Tsychia,R., Gromashevsky,V.L.,
            Usachev,E.V., Mokhonov,V.V., Voronina,A.G., Butenko,A.M.,
            Larichev,V.F., Gubler,D.J. and Lvov,D.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-APR-2003) Molecular Genetic, Ivanovsky Virology
            Institute, Gamalei 16, Moscow 123098, Russia
FEATURES             Location/Qualifiers
     source          1. .10842
                     /organism="West Nile virus"
```

```
                     /mol_type="genomic RNA"
                     /isolate="LEIV-Vlg00-27924"
                     /isolation_source="human blood in 2000"
                     /db_xref="taxon:11082"
                     /country="Russia: Volgograd, low Volga"
     CDS             97. .10398
                     /codon_start=1
                     /product="polyprotein precursor"
                     /protein_id="AAP22087.1"
                     /db_xref="GI:30349732"
                     /translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPV
                     RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
                     KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
                     AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
                     VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
                     VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
                     MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
                     NGCGLFGKGSIDTCAKFACSTKATGRTILKENIKYEVAIFVHGPTTVESHGNYSTQMG
                     ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
                     MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
                     NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
                     PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
                     INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAIHQVFGGAF
                     RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
                     ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
                     QMWESVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
                     FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
                     KIIGTAVKNNLAIHSDLSYWIESRFNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
                     DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
                     TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
                     DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
                     GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYY
                     DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
                     LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
                     TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
                     ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMVCLAISAYTPWAILPSVV
                     GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
                     TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
                     VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNSDVIGLYGNGVIMPNGSYIS
                     AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
                     LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
                     NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
                     TEIPDRAWNSGYEWITEYIGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
                     PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
                     SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
                     PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
                     TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
                     LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
                     VFFLLMQRKGIGKIGLGGVVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
                     QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
                     LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
                     GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
                     VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
                     TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMDKPGLKRGGAKGRTLGE
                     VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
                     RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
                     KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
                     CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
                     RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
                     PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
                     EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
                     EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
                     AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
                     YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
                     TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
                     RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
                     DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
                     QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
                     IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
                     DEKYVDYMSSLKRYEDTILVEDTVL"
ORIGIN

  Query Match          100.0%;  Score 21;  DB 14;  Length 10842;
  Best Local Similarity   100.0%;  Pred. No. 1.6;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCTAAA 21
              |||||||||||||||||||||
Db        129 CCGGGCTGTCAATATGCTAAA 149


RESULT 5
AY277252
LOCUS       AY277252               10845 bp    RNA     linear   VRL 03-MAY-2003
DEFINITION  West Nile virus isolate LEIV-Vlg99-27889, complete genome.
ACCESSION   AY277252
VERSION     AY277252.1  GI:30349727
KEYWORDS    .
SOURCE      West Nile virus (WNV)
  ORGANISM  West Nile virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE   1  (bases 1 to 10845)
  AUTHORS   Prilipov,A.G., Kinney,R.M., Samokhvalov,E.I., Savage,H.M.,
            Alkhovsky,S.V., Tsychia,R., Gromashevsky,V.L., Sadykova,G.K.,
            Shatalov,A.G., Usachev,E.V., Mokhonov,V.V., Voronina,A.G.,
            Butenko,A.M., Larichev,V.F., Gubler,D.J. and Lvov,D.K.
  TITLE     Analysis of a new variants of West Nile virus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 10845)
  AUTHORS   Prilipov,A.G., Kinney,R.M., Samokhvalov,E.I., Savage,H.M.,
            Alkhovsky,S.V., Tsychia,R., Gromashevsky,V.L., Sadykova,G.K.,
            Shatalov,A.G., Usachev,E.V., Mokhonov,V.V., Voronina,A.G.,
            Butenko,A.M., Larichev,V.F., Gubler,D.J. and Lvov,D.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-APR-2003) Molecular Genetic, Ivanovsky Virology
            Institute, Gamalei 16, Moscow 123098, Russia
FEATURES             Location/Qualifiers
     source          1. .10845
                     /organism="West Nile virus"
                     /mol_type="genomic RNA"
                     /isolate="LEIV-Vlg99-27889"
                     /isolation_source="human brain in 1999"
                     /db_xref="taxon:11082"
                     /country="Russia: Volgograd, low Volga"
     CDS             97. .10398
                     /codon_start=1
                     /product="polyprotein precursor"
                     /protein_id="AAP22089.1"
                     /db_xref="GI:30349728"
                     /translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
                     RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
                     KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
                     AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
                     VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
                     VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
                     MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
                     NGCGLFGKGSIDTCAKFACSTKATGRTILKENIKYEVAIFVHGPTTVESHGKYSTQMG
                     ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
                     MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
                     NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
                     PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
                     INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAIHQVFGGAF
                     RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
                     ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
                     QMWESVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
                     FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
                     KIIGTAVKNNLAIHSDLSYWIESRFNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
                     DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
                     TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
                     DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
                     GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYY
                     DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
                     LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
                     TEVMTAVXLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
                     ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMVCLAISAYTPWAILPSVV
```

GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
TEIPDRAWNSGYEWITEYIGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
VFFLLMQRKGIGKIGLGGVVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMDKPGLKRGGAKGRTLGE
VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
EQNQWRSAREAVEDLKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
DEKYVDYMSSLKRYEDTILVEDTVL"

ORIGIN

Query Match 100.0%; Score 21; DB 14; Length 10845;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCGGGCTGTCAATATGCTAAA 21
||||||||||||||||||||
Db 129 CCGGGCTGTCAATATGCTAAA 149

RESULT 6
AF202541
LOCUS AF202541 10945 bp RNA linear VRL 16-DEC-1999
DEFINITION West Nile virus strain HNY1999 polyprotein (C, prM, E, NS1, NS2a, NS2b, NS3, NS4a, NS4b, NS5) gene, complete cds.
ACCESSION AF202541
VERSION AF202541.1 GI:6581069
KEYWORDS .
SOURCE West Nile virus
ORGANISM West Nile virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Flavivirus; Japanese encephalitis virus group.
REFERENCE 1 (bases 1 to 10945)
AUTHORS Jia,X.Y., Briese,T., Jordan,I., Rambaut,A., Chi,H.C., Mackenzie,J.S., Hall,R.A., Scherret,J. and Lipkin,W.I.
TITLE Genetic analysis of West Nile New York 1999 encephalitis virus
JOURNAL Lancet 354 (9194), 1971-1972 (1999)
MEDLINE 20086017
PUBMED 10622305
REFERENCE 2 (bases 1 to 10945)
AUTHORS Jia,X.Y., Briese,T., Jordan,I. and Lipkin,W.I.
TITLE Direct Submission
JOURNAL Submitted (06-NOV-1999) Emerging Diseases Laboratory, Dept. Microbiology & Molecular Genetics and Neurology, University of California, Irvine, 3101 Gillespie Neuroscience Facility, Irvine, CA 92697-4292, USA
FEATURES Location/Qualifiers
source 1. .10945
/organism="West Nile virus"
/mol_type="genomic RNA"
/strain="HNY1999"
/db_xref="taxon:11082"
/country="USA: New York"
/note="isolated from total brain RNA (patient NYC99002) by RT-PCR amplification"
CDS 55. .10356
/codon_start=1
/product="polyprotein"
/protein_id="AAF18443.1"
/db_xref="GI:6581070"
/translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQVG
ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAIN
ITRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
QMWEAVKDELNTPLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVREGNTTECDS
KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMVCLAISAYTPWAILPSVV
GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
TEIPDRAWNSGYEWITEYTGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
VFFLLMQRKGIGKIGLGGAVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
DEKYVDYMSSLKRYEDTTLVEDTVL"
mat_peptide 55. .423

```
                     /product="capsid protein"
                     /note="putative; C"
     mat_peptide     424. .924
                     /product="precursor of M protein"
                     /note="putative; prM"
     mat_peptide     700. .924
                     /product="membrane protein"
                     /note="putative; M"
     mat_peptide     925. .2427
                     /product="envelope protein"
                     /note="putative; E"
     mat_peptide     2428. .3483
                     /product="non-structural protein NS1"
                     /note="putative"
     mat_peptide     3484. .4176
                     /product="non-structural protein NS2a"
                     /note="putative"
     mat_peptide     4177. .4569
                     /product="non-structural protein NS2b"
                     /note="putative"
     mat_peptide     4570. .6426
                     /product="non-structural protein NS3"
                     /note="putative"
     mat_peptide     6427. .6873
                     /product="non-structural protein NS4a"
                     /note="putative"
     mat_peptide     6874. .7638
                     /product="non-structural protein NS4b"
                     /note="putative"
     mat_peptide     7639. .10353
                     /product="non-structural protein NS5"
                     /note="putative"
ORIGIN

  Query Match             100.0%;  Score 21;  DB 14;  Length 10945;
  Best Local Similarity   100.0%;  Pred. No. 1.6;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCTAAA 21
              |||||||||||||||||||||
Db         87 CCGGGCTGTCAATATGCTAAA 107


RESULT 7
WNFCG
LOCUS       WNFCG                  10962 bp ss-RNA     linear   VRL 08-MAY-2002
DEFINITION  West Nile virus RNA, complete genome.
ACCESSION   M12294 M10103
VERSION     M12294.2  GI:11497619
KEYWORDS    .
SOURCE      West Nile virus
  ORGANISM  West Nile virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE   1  (bases 67 to 969)
  AUTHORS   Castle,E., Nowak,T., Leidner,U., Wengler,G. and Wengler,G.
  TITLE     Sequence analysis of the viral core protein and the
            membrane-associated proteins V1 and NV2 of the flavivirus West Nile
            virus and of the genome sequence for these proteins
  JOURNAL   Virology 145 (2), 227-236 (1985)
  MEDLINE   85274372
   PUBMED   2992152
REFERENCE   2  (bases 859 to 2658)
  AUTHORS   Wengler,G., Castle,E., Leidner,U., Nowak,T. and Wengler,G.
  TITLE     Sequence analysis of the membrane protein V3 of the flavivirus West
            Nile virus and of its gene
  JOURNAL   Virology 147 (2), 264-274 (1985)
  MEDLINE   86072082
   PUBMED   3855247
REFERENCE   3  (bases 1 to 10962)
  AUTHORS   Castle,E.
  JOURNAL   Unpublished
REFERENCE   4  (bases 67 to 10485)
  AUTHORS   Castle,E., Leidner,U., Nowak,T., Wengler,G. and Wengler,G.
  TITLE     Primary structure of the West Nile flavivirus genome region coding
            for all nonstructural proteins
  JOURNAL   Virology 149 (1), 10-26 (1986)
  MEDLINE   86124703
   PUBMED   3753811
REFERENCE   5  (bases 1 to 10962)
  AUTHORS   Yamshchikov,V.F., Wengler,G., Perelygin,A.A., Brinton,M.A. and
            Compans,R.W.
  TITLE     An infectious clone of West Nile flavivirus
  JOURNAL   Virology (2000) In press
REFERENCE   6  (bases 1 to 10962)
  AUTHORS   Castle,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1993) Justus-Liebig-Universitat Giessen, Institut
            fur Virologie, 35392, Giessen, Germany
REFERENCE   7  (bases 1 to 10962)
  AUTHORS   Yamshchikov,V.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-DEC-2000) University of Virginia Health Sciences
            Centre, Department of Internal Medicine/GI, Charlottesville, VA
            22906
COMMENT     On Dec 1, 2000 this sequence version replaced gi:336167.
            Draft entry and sequence in computer readable form for
            [1],[2],[4],[3] kindly provided by E.Castle, 12-NOV-1985. The West
            Nile viral genome consists of a 42S viral RNA.  The amino-terminal
            ends of the structural proteins were experimentally determined.  An
            'atg' codon is located at positions 142-144, which could be used
            for an alternative initiation of translation for V2. The
            carboxy-terminal ends of the proteins reported here were not yet
            precisely defined.
FEATURES             Location/Qualifiers
     source          1. .10962
                     /organism="West Nile virus"
                     /virion
                     /mol_type="genomic RNA"
                     /db_xref="taxon:11082"
                     /clone="33/G8; 34/F6"
     CDS             97. .10389
                     /codon_start=1
                     /product="polyprotein precursor"
                     /protein_id="AAA48498.2"
                     /db_xref="GI:11497620"
                     /translation="MSKKPGGPGKNRAVNMLKRGMPRGLSLIGLKRAMLSLIDGKGPI
                     RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSTKQ
                     KKRGGTAGFTILLGLIACAGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
                     AMDVGYLCEDTITYECPVLAAGNDPEDIDCWCTKSSVYVRYGRCTKTRHSRRSRRSLT
                     VQTHGESTLANKKGAWLDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
                     VFAILLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
                     MMNMEAANLADVRSYCYLASVSDLSTRAACPTMGEAHNEKRADPAFVCKQGVVDRGWG
                     NGCGLFGKGSIDTCAKFACTTKATGWIIQKENIKYEVAIFVHGPTTVESHGKIGATQA
                     GRFSITPSAPSYTLKLGEYGEVTVDCEPRSGIDTSAYYVMSVGEKSFLVHREWFMDLN
                     LPWSSAGSTTWRNRETLMEFEEPHATKQSVVALGSQEGALHQALAGAIPVEFSSNTVK
                     LTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFARTPADTGHGTVVLELQYTGTDGPCKV
                     PISSVASLNDLTPVGRLVTVNPFVSVATANSKVLIELEPPFGDSYIVVGRGEQQINHH
                     WHKSGSSIGKAFTTTLRGAQRLAALGDTAWDFGSVGGVFTSVGKAIHQVFGGAFRSLF
                     GGMSWITQGLLGALLLWMGINARDRSIAMTFLAVGGVLLFLSVNVHADTGCAIDIGRQ
                     ELRCGSGVFIHNDVEAWMDRYKFYPETPQGLAKIIQKAHAEGVCGLRSVSRLEHQMWE
                     AIKDELNTLLKENGVDLSVVVEKQNGMYKAAPKRLAATTEKLEMGWKAWGKSIIFAPE
                     LANNTFVIDGPETEECPTANRAWNSMEVEDFGFGLTSTRMFLRIRETNTTECDSKIIG
                     TAVKNNMAVHSDLSYWIESGLNDTWKLERAVLGEVKSCTWPETHTLWGDGVLESDLII
                     PITLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTISDSCEHRGPAARTT
                     TESGKLITDWCCRSCTLPPLRFQTENGCWYGMEIRPTRHDEKTLVQSRVNAYNADMID
                     PFQLGLMVVFLATQEVLRKRWTAKISIPAIMLALLVLVFGGITYTDVLRYVILVGAAF
                     AEANSGGDVVHLALMATFKIQPVFLVASFLKARWTNQESILLMLAAAFFQMAYYDAKN
                     VLSWEVPDVLNSLSVAWMILRAISFTNTSNVVVPLLALLTPGLKCLNLDVYRILLLMV
                     GVGSLIKEKRSSAAKKKGACLICLALASTGVFNPMILAAGLMACDPNRKRGWPATEVM
                     TAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADITWESDA
                     EITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMACLAISAYTPWAILPSVIGFWI
                     TLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFHTLWH
                     TTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGHDEVQMIVVEPGKNVKNV
                     QTKPGVFKTPEGEIGAVTLDYPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYISAIVQ
                     GERMEEPAPAGFEPEMLRKKQITVLDLHPGAGKTRKILPQIIKEAINKRLRTAVLAPT
```

```
                     RVVAAEMSEALRGLPIRYQTSAVHREHSGNEIVDVMCHATLTHRLMSPHRVPNYNLFI
                     MDEAHFTDPASIAARGYIATKVELGEAAAIFMTATPPGTSDPFPESNAPISDMQTEIP
                     DRAWNTGYEWITEYVGKTVWFVPSVKMGNEIALCLQRAGKKVIQLNRKSYETEYPKCK
                     NDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIIEEGDGRVILGEPSAITAASAAQ
                     RRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLVAQLYQPERE
                     KVYTMDGEYRLRGEERKNFLEFLRTADLPVWLAYKVAAAGISYHDRKWCFDGPRTNTI
                     LEDNNEVEVITKLGERKILRPRWADARVYSDHQALKSFKDFASGKRSQIGLVEVLGRM
                     PEHFMVKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIVLIALLSVMSLGVFFL
                     LMQRKGIGKIGLGGVILGAATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEKQRSQ
                     TDNQLAVFLICVLTLVGAVAANEMGWLDKTKNDIGSLLGHRPEARETTLGVESFLLDL
                     RPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDVGVS
                     ALLLAVGCWGQVTLTVTVTAAALLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNVVVD
                     GIVATDVPELERTTPVMQKKVGQIILILVSMAAVVVNPSVRTVREAGILTTAAAVTLW
                     ENGASSVWNATTAIGLCHIMRGGWLSCLSIMWTLIKNMEKPGLKRGGAKGRTLGEVWK
                     ERLNHMTKEEFTRYRKEAITEVDRSAAKHARREGNITGGHPVSRGTAKLRWLVERRFL
                     EPVGKVVDLGCGRGGWCYYMATQKRVQEVKGYTKGGPGHEEPQLVQSYGWNIVTMKSG
                     VDVFYRPSEASDTLLCDIGESSSSAEVEEHRTVRVLEMVEDWLHRGPKEFCIKVLCPY
                     MPKVIEKMETLQRRYGGGLIRNPLSRNSTHEMYWVSHASGNIVHSVNMTSQVLLGRME
                     KKTWKGPQFEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLKKEYSSTWHQDANHPYR
                     TWNYHGSYEVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFKEKV
                     DTKAPEPPEGVKYVLNETTNWLWAFLARDKKPRMCSREEFIGKVNSNAALGAMFEEQN
                     QWKNAREAVEDPKFWEMVDEEREAHLRGECNTCIYNMMGKREKKPGEFGKAKGSRAIW
                     FMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILKEVGTKPGGKVYAD
                     DTAGWDTRITKADLENEAKVLELLDGEHRRLARSIIELTYRHKVVKVMRPAADGKTVM
                     DVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLGKGKGPKVRTW
                     LFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWYDWQ
                     QVPFCSNHFTELIMKDGRTLVVPCRGQDELIGRARISPGAGWNVRDTACLAKSYAQMW
                     LLLYFHRRDLRLMANAICSAVPANWVPTGRTTWSIHAKGEWMTTEDMLAVWNRVWIEE
                     NEWMEDKTPVERWSDVPYSGKREDIWCGSLIGTRTRATWAENIHVAINQVRSVIGEEK
                     YVDYMSSLRRYEDTIVVEDTVL"
     mat_peptide     97. .372
                     /product="V2 (14kd core protein)"
     sig_peptide     409. .465
                     /note="NV2 signal peptide"
     mat_peptide     466. .765
                     /product="NV2 (20.5 kd membrane-associated glycoprotein)"
     mat_peptide     742. .765
                     /product="V1 (7 kd membrane-associated nonglycosylated
                     protein"
     sig_peptide     919. .966
                     /note="V3 signal peptide"
     mat_peptide     967. .2457
                     /product="V3 (50 kd membrane-associated glycoprotein;
                     putative); putative"
     sig_peptide     2386. .2457
                     /note="nonstructural protein NV4 signal peptide"
     mat_peptide     2458. .6426
                     /product="nonstructural protein NV4"
     mat_peptide     7834. .10380
                     /product="nonstructural protein NV5"
ORIGIN

  Query Match           100.0%;  Score 21;  DB 14;  Length 10962;
  Best Local Similarity  100.0%;  Pred. No. 1.6;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCTAAA 21
              |||||||||||||||||||||
Db        129 CCGGGCTGTCAATATGCTAAA 149


RESULT 8
AF317203
LOCUS       AF317203               10972 bp    RNA     linear   VRL 11-FEB-2001
DEFINITION  West Nile virus VLG-4 polyprotein precursor, gene, complete cds.
ACCESSION   AF317203
VERSION     AF317203.1  GI:12744408
KEYWORDS    .
SOURCE      West Nile virus
  ORGANISM  West Nile virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE   1  (bases 1 to 10972)
  AUTHORS   Platonov,A.E., Karan,L., Yazishina,S., Obukhov,I.L., Shipulina,O.
            and Shipulin,G.A.
  TITLE     Genetic similarity of West Nile viruses caused epidemics in
            Volgograd 1999 and Romania 1996
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 10972)
  AUTHORS   Karan,L., Yazishina,S., Obukhov,I.L., Shipulina,O., Shipulin,G.A.
            and Platonov,A.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-2000) Central Research Institute of Epidemiology,
            Novogireevskaya Str. 3A, Moscow 111123, Russia
FEATURES             Location/Qualifiers
     source          1. .10972
                     /organism="West Nile virus"
                     /mol_type="genomic RNA"
                     /isolate="VLG-4"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:11082"
                     /country="Russia: Volgograd"
                     /note="isolated from brain of patient that died of
                     encephalitis in September 1999"
     CDS             65. .10366
                     /codon_start=1
                     /product="polyprotein precursor"
                     /protein_id="AAK06624.1"
                     /db_xref="GI:12744409"
                     /translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
                     RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
                     KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
                     AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
                     VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
                     VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
                     MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
                     NGCGLFGKGSIDTCAKFACSTKATGRTILKENIKYEVAIFVHGPTTVESHGKYSTQMG
                     ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
                     MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
                     NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
                     PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
                     INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAIHQVFGGAF
                     RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
                     ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
                     QMWESVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
                     FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
                     KIIGTAVKNNLAIHSDLSYWIESRFNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
                     DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
                     TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
                     DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
                     GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYY
                     DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
                     LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
                     TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
                     ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMVCLAISAYTPWAILPSVV
                     GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
                     TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
                     VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
                     AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
                     LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
                     NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
                     TEIPDRAWNSGYEWITEYIGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
                     PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
                     SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
                     PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
                     TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
                     LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
                     VFFLLMQRKGIGKIGLGGVVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
                     QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
                     LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
                     GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
                     VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
                     TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMDKPGLKRGGAKGRTLGE
                     VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
                     RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
                     KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
```

```
                     CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
                     RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
                     PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
                     EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
                     EQNQWRSAREAVEDLKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
                     AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
                     YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
                     TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
                     RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
                     DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
                     QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
                     IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
                     DEKYVDYMSSLKRYEDTILVEDTVL"
     mat_peptide     65. .433
                     /product="putative nucleocapsid protein C"
     mat_peptide     434. .709
                     /product="putative pre-membrane protein prM"
     mat_peptide     710. .934
                     /product="putative membrane protein M"
     mat_peptide     935. .2437
                     /product="putative envelope glycoprotein E"
     mat_peptide     2438. .3493
                     /product="putative non-structural protein NS1"
     mat_peptide     3494. .4186
                     /product="putative non-structural protein NS2A"
     mat_peptide     4187. .4579
                     /product="putative non-structural protein NS2B"
     mat_peptide     4580. .6436
                     /product="putative non-structural protein NS3"
     mat_peptide     6437. .6883
                     /product="putative non-structural protein NS4A"
     mat_peptide     6884. .7648
                     /product="putative non-structural protein NS4B"
     mat_peptide     7649. .10363
                     /product="putative non-structural protein NS5"
ORIGIN

  Query Match           100.0%;  Score 21;  DB 14;  Length 10972;
  Best Local Similarity  100.0%;  Pred. No. 1.6;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCTAAA 21
              |||||||||||||||||||||
Db         97 CCGGGCTGTCAATATGCTAAA 117


RESULT 9
AF206518
LOCUS       AF206518               10975 bp    DNA     linear   VRL 08-MAY-2000
DEFINITION  West Nile virus isolate 2741, complete genome.
ACCESSION   AF206518
VERSION     AF206518.2  GI:7717200
KEYWORDS    .
SOURCE      West Nile virus
  ORGANISM  West Nile virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE   1  (bases 1 to 10975)
  AUTHORS   Anderson,J.F., Andreadis,T.G., Vossbrinck,C.R., Tirrell,S.,
            Wakem,E.M., French,R.A., Garmendia,A.E. and Van Kruiningen,H.J.
  TITLE     Isolation of West Nile virus from mosquitoes, crows, and a Cooper's
            hawk in Connecticut
  JOURNAL   Science 286 (5448), 2331-2333 (1999)
  MEDLINE   20070287
   PUBMED   10600741
REFERENCE   2  (bases 1 to 10975)
  AUTHORS   Vossbrinck,C.R., Anderson,J.F. and Andreadis,T.G.
  TITLE     Genome Sequence of West Nile Virus from Culex pipiens isolate
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 10975)
  AUTHORS   Anderson,J.F., Andreadis,T.G. and Vossbrinck,C.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-1999) Soil and Water, Connecticut Agricultural
            Experiment Station, 123 Huntington Street, New Haven, CT 06511, USA
REFERENCE   4  (bases 1 to 10975)
  AUTHORS   Anderson,J.F., Andreadis,T.G. and Vossbrinck,C.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAY-2000) Soil and Water, Connecticut Agricultural
            Experiment Station, 123 Huntington Street, New Haven, CT 06511, USA
  REMARK    Sequence update by submitter
COMMENT     On May 8, 2000 this sequence version replaced gi:6636507.
FEATURES             Location/Qualifiers
     source          1. .10975
                     /organism="West Nile virus"
                     /mol_type="genomic DNA"
                     /strain="Connecticut 1999"
                     /isolate="2741"
                     /specific_host="Culex pipiens"
                     /db_xref="taxon:11082"
                     /country="USA:Connecticut, Greenwich-Stanford town line"
     CDS             79. .10380
                     /codon_start=1
                     /product="polyprotein precursor"
                     /protein_id="AAF20205.2"
                     /db_xref="GI:7717201"
                     /translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
                     RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
                     KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
                     AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
                     VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
                     VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
                     MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
                     NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQVG
                     ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
                     MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
                     NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
                     PCKVPISSVASLNDLTPVGRLVTVNPFVSMATANAKVLIELEPPFGDSYIVVGRGEQQ
                     INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
                     RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
                     ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
                     QMWEAVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
                     FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
                     KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
                     DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
                     TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
                     DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
                     GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
                     DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
                     LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
                     TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
                     ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMVCLAISAYTPWAILPSVV
                     GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
                     TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
                     VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
                     AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
                     LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
                     NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
                     TEIPDRAWNSGYEWITEYTGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
                     PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEARVILGEPSAVTAA
                     SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
                     PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
                     TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
                     LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
                     VFFLLMQRKGIGKIGLGGAVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
                     QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
                     LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
                     GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
                     VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
                     TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
                     VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
                     RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
                     KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
                     CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
                     RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
                     PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
                     EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
```

EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
DEKYVDYMSSLKRYEDTTLVEDTVL"
ORIGIN

Query Match 100.0%; Score 21; DB 14; Length 10975;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCGGGCTGTCAATATGCTAAA 21
||||||||||||||||||||
Db 111 CCGGGCTGTCAATATGCTAAA 131

RESULT 10
AY262283
LOCUS AY262283 10984 bp RNA linear VRL 29-OCT-2003
DEFINITION West Nile virus isolate KN3829 polyprotein gene, complete cds.
ACCESSION AY262283
VERSION AY262283.1 GI:30230630
KEYWORDS .
SOURCE West Nile virus (WNV)
ORGANISM West Nile virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Flavivirus; Japanese encephalitis virus group.
REFERENCE 1 (bases 1 to 10984)
AUTHORS Charrel,R.N., Brault,A.C., Gallian,P., Lemasson,J.-J., Murgue,B., Murri,S., Pastorino,B., Zeller,H., de Chesse,R., de Micco,P. and de Lamballerie,X.
TITLE Evolutionary relationship between Old World West Nile virus strains. Evidence for viral gene flow between africa, the middle east, and europe
JOURNAL Virology 315 (2), 381-388 (2003)
MEDLINE 22949215
PUBMED 14585341
REFERENCE 2 (bases 1 to 10984)
AUTHORS Brault,A.C. and de Lamballerie,X.
TITLE Direct Submission
JOURNAL Submitted (25-MAR-2003) Division of Vector-Borne Infectious Diseases, Centers for Disease Control and Prevention, P.O. Box 2087, Fort Collins, CO 80522, USA
FEATURES Location/Qualifiers
source 1. .10984
/organism="West Nile virus"
/mol_type="genomic RNA"
/isolate="KN3829"
/specific_host="Culex univittatus"
/db_xref="taxon:11082"
5'UTR 1. .60
CDS 61. .10362
/codon_start=1
/product="polyprotein"
/protein_id="AAP20887.1"
/db_xref="GI:30230631"
/translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
KKRGGNTGIAAMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
NGCGLFGKGSIDTCAKFACSTKATGRTILKENIKYEVAIFVHGPTTVESHGNYSTQIG
ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
QMWESVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYADVLRYVILV
GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMACLAISAYTPWAILPSVV
GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
LAPTRVVAAEMAEALRGLPIRYQTSAVTREHNGNEIVDVMCHATLTHRLMSPHRVPNY
NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
TEIPDRAWNSGYEWITEYIGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
VFFLLMQRKGIGKIGLGGVVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMDKPGLKRGGAKGRTLGE
VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
DEKYVDYMSSLKRYEDTTLVEDTVL"
mat_peptide 61. .429
/product="capsid"
mat_peptide 430. .705
/product="prM"
mat_peptide 706. .930
/product="M"
mat_peptide 931. .2433
/product="envelope"
mat_peptide 2434. .3492
/product="NS1"
mat_peptide 3493. .4182
/product="NS2A"
mat_peptide 4183. .4575
/product="NS2B"
mat_peptide 4576. .6432
/product="NS3"
mat_peptide 6433. .6879
/product="NS4A"
mat_peptide 6880. .7644
/product="NS4B"
mat_peptide 7645. .10359
/product="NS5"
3'UTR 10363. .10984
ORIGIN

Query Match 100.0%; Score 21; DB 14; Length 10984;

```
  Best Local Similarity   100.0%;  Pred. No. 1.6;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCTAAA 21
              |||||||||||||||||||||
Db         93 CCGGGCTGTCAATATGCTAAA 113


RESULT 11
AY268132
LOCUS       AY268132               10989 bp    RNA     linear   VRL 03-NOV-2003
DEFINITION  West Nile virus strain PaAn001 polyprotein (pol) gene, complete
            cds.
ACCESSION   AY268132
VERSION     AY268132.1  GI:33242574
KEYWORDS    .
SOURCE      West Nile virus (WNV)
  ORGANISM  West Nile virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE   1  (bases 1 to 10989)
  AUTHORS   Charrel,R.N., Brault,A.C., Gallian,P., Lemasson,J.-J., Murgue,B.,
            Murri,S., Pastorino,B., Zeller,H., de chesse,R., de Micco,P. and de
            Lamballerie,X.
  TITLE     Evolutionary relationship between Old World West Nile virus
            strains. Evidence for viral gene flow between africa, the middle
            east, and europe
  JOURNAL   Virology 315 (2), 381-388 (2003)
  MEDLINE   22949215
   PUBMED   14585341
REFERENCE   2  (bases 1 to 10989)
  AUTHORS   de Lamballerie,X., Brault,A.C., Gallian,P., Lemasson,J., Murgue,B.,
            Murri,S., Pastorino,B., Zeller,H., Dechesse,R., de Micco,P. and
            Charrel,R.N.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2003) Virology, Medical University, 27 bd Jean
            Moulin, Marseille 13005, France
FEATURES             Location/Qualifiers
     source          1. .10989
                     /organism="West Nile virus"
                     /mol_type="genomic RNA"
                     /strain="PaAn001"
                     /db_xref="taxon:11082"
     gene            77. .10378
                     /gene="pol"
     CDS             77. .10378
                     /gene="pol"
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="AAQ00998.1"
                     /db_xref="GI:33242575"
                     /translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
                     RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
                     KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
                     AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
                     VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
                     VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
                     MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
                     NGCGLFGKGSIDTCAKFACSTKATGRTILKENIKYEVAIFVHGPTTVESHGNYSTQIG
                     ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
                     MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
                     NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
                     PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
                     INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
                     RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
                     ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKVIQKAHKEGVCGLRSVSRLEH
                     QMWESVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
                     FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
                     KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
                     DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
                     TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
                     DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
                     GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
                     DARQILLWGIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
                     LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
                     TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
                     ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMACLAISAYTPWAILPSVV
                     GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
                     TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
                     VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
                     AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
                     LAPTRVVAAEMAEALRGLPIRYHTSAVTREHNGNEIVDVMCHATLTHRLMSPHRVPNY
                     NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
                     TEIPDRAWNSGYEWITEYIGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
                     PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
                     SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPSGLIAQFYQ
                     PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
                     TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
                     LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
                     VFFLLMQRKGIGKIGLGGVVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
                     QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
                     LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
                     GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
                     VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
                     TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMDKPGLKRGGAKGRTLGE
                     VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
                     RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
                     KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
                     CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
                     RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
                     PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
                     EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
                     EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
                     AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
                     YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
                     TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
                     RTWLFENGEERLSRMAVSGDDCVVKPLDDRFAISLHFLNAMSKVRKDIQEWKPSTGWY
                     DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
                     QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
                     IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
                     DEKYVDYMSSLKRYEDTTLVEDTVL"
ORIGIN

  Query Match             100.0%;  Score 21;  DB 14;  Length 10989;
  Best Local Similarity   100.0%;  Pred. No. 1.6;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCTAAA 21
              |||||||||||||||||||||
Db        109 CCGGGCTGTCAATATGCTAAA 129


RESULT 12
AY268133
LOCUS       AY268133               10989 bp    RNA     linear   VRL 03-NOV-2003
DEFINITION  West Nile virus strain PaH001 polyprotein (pol) gene, complete cds.
ACCESSION   AY268133
VERSION     AY268133.1  GI:33242576
KEYWORDS    .
SOURCE      West Nile virus (WNV)
  ORGANISM  West Nile virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE   1  (bases 1 to 10989)
  AUTHORS   Charrel,R.N., Brault,A.C., Gallian,P., Lemasson,J.-J., Murgue,B.,
            Murri,S., Pastorino,B., Zeller,H., de chesse,R., de Micco,P. and de
            Lamballerie,X.
  TITLE     Evolutionary relationship between Old World West Nile virus
            strains. Evidence for viral gene flow between africa, the middle
            east, and europe
  JOURNAL   Virology 315 (2), 381-388 (2003)
  MEDLINE   22949215
   PUBMED   14585341
REFERENCE   2  (bases 1 to 10989)
  AUTHORS   de Lamballerie,X., Brault,A.C., Gallian,P., Lemasson,J., Murgue,B.,
            Murri,S., Pastorino,B., Zeller,H., Dechesse,R., de Micco,P. and
```

```
          Charrel,R.N.
  TITLE   Direct Submission
  JOURNAL Submitted (03-APR-2003) Virology, Medical University, 27 bd Jean
          Moulin, Marseille 13005, France
FEATURES             Location/Qualifiers
     source          1. .10989
                     /organism="West Nile virus"
                     /mol_type="genomic RNA"
                     /strain="PaH001"
                     /db_xref="taxon:11082"
     gene            77. .10378
                     /gene="pol"
     CDS             77. .10378
                     /gene="pol"
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="AAQ00999.1"
                     /db_xref="GI:33242577"
                     /translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
                     RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
                     KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
                     AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
                     VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
                     VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
                     MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
                     NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQIG
                     ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
                     MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGTIPVEFSS
                     NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
                     PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
                     INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
                     RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
                     ISRQELRCGSGVFIHNDVKAWMDRYKYYLETPQGLAKIIQKAHKEGVCGLRSVSRLEH
                     QMWEAVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
                     FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
                     KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
                     DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
                     TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
                     DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
                     GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
                     DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
                     LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
                     TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
                     ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMACLAISAYTPWAILPSVV
                     GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
                     TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
                     VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
                     AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
                     LAPTRVVAAEMAEALRGLPIRYQTSAVTREHNGNEIVDVMCHATLTHRLMSPHRVPNY
                     NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
                     TEIPDRAWNSGYEWITEYIGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
                     PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
                     SAAQRRGRVGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
                     PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
                     TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
                     LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
                     VFFLLMQRKGIGKIGLGGAVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
                     QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
                     LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
                     GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
                     VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
                     TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
                     VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
                     RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
                     KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
                     CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
                     RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
                     PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
                     EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
                     EQNQWRSAREAVEDPKFWEMVDEERE AHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
                     AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
                     YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
                     TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
```

```
                     RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
                     DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
                     QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
                     IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
                     DEKYVDYMSSLKRYEDTTLVEDTVL"
ORIGIN

  Query Match             100.0%;  Score 21;  DB 14;  Length 10989;
  Best Local Similarity   100.0%;  Pred. No. 1.6;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCTAAA 21
              |||||||||||||||||||||
Db        109 CCGGGCTGTCAATATGCTAAA 129


RESULT 13
AY274504
LOCUS       AY274504               11022 bp    mRNA    linear   VRL 02-JUL-2003
DEFINITION  Kunjin virus clone FLSDX polyprotein mRNA, complete cds.
ACCESSION   AY274504
VERSION     AY274504.1  GI:32306849
KEYWORDS    .
SOURCE      Kunjin virus
  ORGANISM  Kunjin virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE   1  (bases 1 to 11022)
  AUTHORS   Liu,W.J., Chen,H.B. and Khromykh,A.A.
  TITLE     Molecular and Functional Analyses of Kunjin Virus Infectious cDNA
            Clones Demonstrate the Essential Roles for NS2A in Virus Assembly
            and for a Nonconservative Residue in NS3 in RNA Replication
  JOURNAL   J. Virol. 77 (14), 7804-7813 (2003)
  MEDLINE   22713678
   PUBMED   12829820
REFERENCE   2  (bases 1 to 11022)
  AUTHORS   Khromykh,A.A., Liu,W.J. and Chen,H.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2003) Clinical Medical Virology Centre,
            University of Queensland/Sir Albert Sakzewski Virus Research
            Centre, Royal Children's Hospital, Herston Rd., Herston, Brisbane,
            QLD 4029, Australia
FEATURES             Location/Qualifiers
     source          1. .11022
                     /organism="Kunjin virus"
                     /mol_type="mRNA"
                     /db_xref="taxon:11077"
                     /clone="FLSDX"
     5'UTR           1. .96
     CDS             97. .10398
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="AAP78941.1"
                     /db_xref="GI:32306850"
                     /translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLTGLKRAMLSLIDGRGPT
                     RFVLALLAFFRFTAIAPTRAVLDRWRSVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
                     KKRGGKTGIAFMIGLIAGVGAVTLSNFQGKVMMTVNATDVTDIITIPTAAGKNLCIVR
                     AMDVGHMCDDTITYECPVLSAGNDPEDIDCWCTKLAVYVRYGRCTKTRHSRRSRRSLT
                     VQTHGESTLSNKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
                     VFAVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
                     MMNMEAANLAEVRSYCYLATVSELSTKAACPTMGEAHNDKRADPSFVCKQGVVDRGWG
                     NGCGLFGKGSIDTCAKFACSTKATGRTILKENIKYEVAIFVHGPTTVESHGNYFTQTG
                     AAQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTSAYYVMTVGTKTFLVHREWF
                     MDLNLPWSSAESNVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
                     NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFRFLGTPADTGHGTVVLELQYTGTDG
                     PCKIPISSVASLNDLTPVGRLVTVNPFVSVSTANAKVLIELEPPFGDSYIVVGRGEQQ
                     INHHWHKSGSSIGKAFTATLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
                     RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
                     ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
                     QMWEAVKDELNTLLKENGVDLSIVVEKQEGMYKSAPRRLTATTEKLEIGWKAWGKSIL
                     FAPELANNTFVIDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLRVRESNTTECDS
                     KIIGTAVKNNLAIHSDLSYWIESRFNDTWKLERAVLGEVKSCTWPETHTLWGDGVLES
                     DLIIPITLAGLRSNHNRRPGYKTQSQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
```

```
                TRTTTESGKLITDWCCRSCTLPPLRYQTDNGCWYGMEIRPQRHDEKTLVQSQVNAYNA
                DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
                GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAAFFQMAYY
                DARQILLWEMPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
                LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGFFNPMILAAGLVACDPNRKRGWPA
                TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
                EGDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMACLAISAYTPWAILPSVV
                GFWITLQYTKRGGVLWDTPSPKEYKRGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
                TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
                VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
                AIVQGERMDEPVPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
                LAPTRVVAAEMAEALRGLPIRYQTSAVAREHNGNEIVDVMCHATLTHRLMSPHRVPNY
                NLFVMDEAHFTDPASIAARGYISTRVELGEAAAIFMTATPPGTSDPFPESNAPISDLQ
                TEIPDRAWNSGYEWITEYIGKTVWFVPSVKMGNEIALCLQRAGKKVIQLNRKSYETEY
                PKCKNDDWDFVVTTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
                SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNCAHWTEARIMLDNINMPNGLIAQFYQ
                PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
                TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKSFKDFASGKRSQIGFIEV
                LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
                VFFLLMQRKGIGKIGLGGVVLGAATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
                QRSQTDNQLAVFLICVLTLVGAVAANEMGWLDKTKSDISGLFGQRIETKENFSIGEFL
                LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
                GVSALLLAAGCWGQVTLTVTVTSATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
                VVDGIVATDVPELERTTPIMQKKVGQVMLILVSLAALVVNPSVKTVREAGILITAAAV
                TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLVKNMEKPGLKRGGAKGRTLGE
                VWKERLNQMTKEEFIRYRKEAITEVDRSAAKHARKERNITGGHPVSRGTAKLRWLVER
                RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
                KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTLRVLEMVEDWLHRGPKEFCVKVL
                CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
                RMEKKTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
                PYRTWNYHGSYEVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
                EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
                EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
                AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGRI
                YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
                TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
                RTWLSENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
                DWQQVPFCSNHFTELIMKDGRTLVTPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
                QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
                IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRSIIG
                DEKYVDYMSSLKRYEDTTLVEDTVL"
mat_peptide     97. .411
                /product="core protein"
                /note="C"
mat_peptide     412. .966
                /product="premembrane protein"
                /note="prM"
mat_peptide     967. .2469
                /product="envelope protein"
                /note="E"
mat_peptide     2470. .3525
                /product="nonstructural protein 1"
                /note="NS1"
mat_peptide     3526. .4218
                /product="nonstructural protein 2A"
                /note="NS2A"
mat_peptide     4219. .4611
                /product="nonstructural protein 2B"
                /note="NS2B"
mat_peptide     4612. .6468
                /product="nonstructural protein 3"
                /note="serine protease and helicase; NS3"
mat_peptide     6469. .6915
                /product="nonstructural protein 4A"
                /note="NS4A"
mat_peptide     6916. .7680
                /product="nonstructural protein 4B"
                /note="NS4B"
mat_peptide     7681. .10395
                /product="nonstructural protein 5"
                /note="methyl transferase and RNA-dependent RNA
                polymerase; NS5"
3'UTR           10399. .11022

ORIGIN

  Query Match           100.0%;  Score 21;  DB 14;  Length 11022;
  Best Local Similarity  100.0%;  Pred. No. 1.6;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCTAAA 21
              |||||||||||||||||||||
Db        129 CCGGGCTGTCAATATGCTAAA 149


RESULT 14
AY274505
LOCUS       AY274505               11022 bp    mRNA    linear   VRL 02-JUL-2003
DEFINITION  Kunjin virus clone pAKUN polyprotein mRNA, complete cds.
ACCESSION   AY274505
VERSION     AY274505.1  GI:32306851
KEYWORDS    .
SOURCE      Kunjin virus
  ORGANISM  Kunjin virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE   1  (bases 1 to 11022)
  AUTHORS   Liu,W.J., Chen,H.B. and Khromykh,A.A.
  TITLE     Molecular and Functional Analyses of Kunjin Virus Infectious cDNA
            Clones Demonstrate the Essential Roles for NS2A in Virus Assembly
            and for a Nonconservative Residue in NS3 in RNA Replication
  JOURNAL   J. Virol. 77 (14), 7804-7813 (2003)
  MEDLINE   22713678
   PUBMED   12829820
REFERENCE   2  (bases 1 to 11022)
  AUTHORS   Khromykh,A.A., Liu,W.J. and Chen,H.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2003) Clinical Medical Virology Centre,
            University of Queensland/Sir Albert Sakzewski Virus Research
            Centre, Royal Children's Hospital, Herston Rd., Herston, Brisbane,
            QLD 4029, Australia
FEATURES             Location/Qualifiers
     source          1. .11022
                     /organism="Kunjin virus"
                     /mol_type="mRNA"
                     /db_xref="taxon:11077"
                     /clone="pAKUN"
     5'UTR           1. .96
     CDS             97. .10398
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="AAP78942.1"
                     /db_xref="GI:32306852"
                     /translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLTGLKRAMLSLIDGRGPT
                     RFVLALLAFFRFTAIAPTRAVLDRWRSVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
                     KKRGGKTGIAFMIGLIAGVGAVTLSNFQGKVMMTVNATDVTDIITIPTAAGKNLCIVR
                     AMDVGHMCDDTITYECPVLSAGNDPEDIDCWCTKLAVYVRYGRCTKTRHSRRSRRSLT
                     VQTHGESTLSNKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
                     VFAVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
                     MMNMEAANLAEVRSYCYLATVSELSTKAACPTMGEAHNDKRADPSFVCKQGVVDRGWG
                     NGCGLFGKGSIDTCAKFACSTKATGRTILKENIKYEVAIFVHGPTTVESHGNYFTQTG
                     AAQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTSAYYVMTVGTKTFLVHREWF
                     MDLNLPWSSAESNVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
                     NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFRFLGTPADTGHGTVVLELQYTGTDG
                     PCKIPISSVASLNDLTPVGRLVTVNPFVSVSTANAKVLIELEPPFGDSYIVVGRGEQQ
                     INHHWHKSGSSIGKAFTATLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
                     RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
                     ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
                     QMWEAVKDELNTLLKENGVDLSIVVEKQEGMYKSAPRRLTATTEKLEIGWKAWGKSIL
                     FAPELANNTFVIDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLRVRESNTTECDS
                     KIIGTAVKNNLAIHSDLSYWIESRFNDTWKLERAVLGEVKSCTWPETHTLWGDGVLES
                     DLIIPITLAGLRSNHNRRPGYKTQSQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
                     TRTTTESGKLITDWCCRSCTLPPLRYQTDNGCWYGMEIRPQRHDEKTLVQSQVNAYNA
                     DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVNLV
                     GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAAFFQMAYY
                     DARQILLWEMPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYKIL
                     LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGFFNPMILAAGLVACDPNRKRGWPA
```

```
                     TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
                     EGDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMACLAISAYTPWAILPSVV
                     GFWITLQYTKRGGVLWDTPSPKEYKRGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
                     TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
                     VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
                     AIVQGERMDEPVPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
                     LAPTRVVAAEMAEALRGLPIRYQTSAVAREHNGNEIVDVMCHATLTHRLMSPHRVPNY
                     NLFVMDEAHFTDPASIAARGYISTRVELGEAAAIFMTATPPGTSDPFPESNAPISDLQ
                     TEIPDRAWNSGYEWITEYIGKTVWFVPSVKMGNEIALCLQRAGKKVIQLNRKSYETEY
                     PKCKNDDWDFVVTTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
                     SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNCAHWTEARIMLDNINMPNGLIAQFYQ
                     PEREKVHTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVPYHDRRWCFDGPR
                     TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKSFKDFASGKRSQIGFIEV
                     LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
                     VFFLLMQRKGIGKIGLGGVVLGAATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
                     QRSQTDNQLAVFLICVLTLVGAVAANEMGWLDKTKSDISGLFGQRIETKENFSIGEFL
                     LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
                     GVSALLLAAGCWGQVTLTVTVTSATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
                     VVDGIVATDVPELERTTPIMQKKVGQVMLILVSLAALVVNPSVKTVREAGILITAAAV
                     TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLVKNMEKPGLKRGGAKGRTLGE
                     VWKERLNQMTKEEFIRYRKEAITEVDRSAAKHARKERNITGGHPVSRGTAKLRWLVER
                     RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
                     KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTLRVLEMVEDWLHRGPKEFCVKVL
                     CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
                     RMEKKTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
                     PYRTWNYHGSYEVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
                     EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
                     EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
                     AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGRI
                     YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
                     TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
                     RTWLSENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
                     DWQQVPFCSNHFTELIMKDGRTLVTPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
                     QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
                     IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRSIIG
                     DEKYVDYMSSLKRYEDTTLVEDTVL"
     mat_peptide     97. .411
                     /product="core protein"
                     /note="C"
     mat_peptide     412. .966
                     /product="premembrane protein"
                     /note="prM"
     mat_peptide     967. .2469
                     /product="envelope protein"
                     /note="E"
     mat_peptide     2470. .3525
                     /product="nonstructural protein 1"
                     /note="NS1"
     mat_peptide     3526. .4218
                     /product="nonstructural protein 2A"
                     /note="NS2A"
     mat_peptide     4219. .4611
                     /product="nonstructural protein 2B"
                     /note="NS2B"
     mat_peptide     4612. .6468
                     /product="nonstructural protein 3"
                     /note="serine protease and helicase; NS3"
     mat_peptide     6469. .6915
                     /product="nonstructural protein 4A"
                     /note="NS4A"
     mat_peptide     6916. .7680
                     /product="nonstructural protein 4B"
                     /note="NS4B"
     mat_peptide     7681. .10395
                     /product="nonstructural protein 5"
                     /note="methyl transferase and RNA-dependent RNA
                     polymerase; NS5"
     3'UTR           10399. .11022
ORIGIN

  Query Match           100.0%;  Score 21;  DB 14;  Length 11022;
  Best Local Similarity  100.0%;  Pred. No. 1.6;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCTAAA 21
              |||||||||||||||||||||
Db        129 CCGGGCTGTCAATATGCTAAA 149


RESULT 15
AY490240
LOCUS       AY490240               11028 bp    RNA     linear   VRL 08-APR-2004
DEFINITION  West Nile virus strain Chin-01, complete genome.
ACCESSION   AY490240
VERSION     AY490240.2  GI:46277828
KEYWORDS    .
SOURCE      West Nile virus (WNV)
  ORGANISM  West Nile virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE   1  (bases 1 to 11028)
  AUTHORS   Jiang,T., Qin,E. and Deng,Y.
  TITLE     Sequence determination and analysis of West Nile Virus Chin strain
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 11028)
  AUTHORS   Jiang,T., Qin,E. and Deng,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-NOV-2003) Virology, Institute of Microbiology and
            Epidemiology, FengTai Dongda Street, Beijing 100071, China
REFERENCE   3  (bases 1 to 11028)
  AUTHORS   Jiang,T., Qin,E. and Deng,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-2004) Virology, Institute of Microbiology and
            Epidemiology, FengTai Dongda Street, Beijing 100071, China
  REMARK    Sequence update by submitter
COMMENT     On Apr 8, 2004 this sequence version replaced gi:40362614.
FEATURES             Location/Qualifiers
     source          1. .11028
                     /organism="West Nile virus"
                     /mol_type="genomic RNA"
                     /strain="Chin-01"
                     /db_xref="taxon:11082"
                     /country="China"
     CDS             97. .10398
                     /codon_start=1
                     /product="polyprotein precursor"
                     /protein_id="AAR84614.1"
                     /db_xref="GI:40362615"
                     /translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
                     RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKRELGTLTSAINRRSSKQ
                     KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
                     AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
                     VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
                     VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
                     MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCKQGVVDRGWG
                     NGCGLFGKGSIDTCAKFACSTKATGRTILKENIKYEVAIFVHGPTTVESHGNYPTQIG
                     ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
                     MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
                     NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
                     PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
                     INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
                     RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAIGGVLLFLSVNVHADTGCAID
                     ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
                     QMWEAVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
                     FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
                     KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGEGILES
                     DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
                     TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDGKTLVQSQVNAYNA
                     DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
                     GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
                     DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
                     LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
                     TEVMTAAGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
                     ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMACLAISAYTPWAILPSVV
                     GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
                     TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQCKWNGQDEVQMIVVEPGKN
```

VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
LAPTRVVAAEMAEALRGLPIRYQTSAVTREHNGNEIVDVMCHATLTHRLMSPHRVPNY
NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
TEIPDRAWNSGYEWITEYIGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
VFFLLMQRKGIGKIGLGGVVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
QRSQTDNQLAVFLICVLTLVSAVAANEMGWLDKTKNDISSLFGQRIDVKENFSMGEFL
LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPKEFCVKVL
CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIKKVNSNAALGAMFE
EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPEDVEKLTKGKGPKV
RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEGMLEVWNRVW
IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRTRATWAENIQVAINQVRAIIG
DEKYVDYMSSLKRYEDTTLVEDTVL"

ORIGIN

Query Match 100.0%; Score 21; DB 14; Length 11028;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCGGGCTGTCAATATGCTAAA 21
||||||||||||||||||||
Db 129 CCGGGCTGTCAATATGCTAAA 149

RESULT 16
AX576542
LOCUS AX576542 11029 bp DNA linear PAT 08-JAN-2003
DEFINITION Sequence 1 from Patent WO02081511.
ACCESSION AX576542
VERSION AX576542.1 GI:27646162
KEYWORDS .
SOURCE Flavivirus sp.
ORGANISM Flavivirus sp.
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Flavivirus.
REFERENCE 1
AUTHORS Despres,P., Deubel,V., Guenet,J.L., Drouet,M.T., Malkinson,M.K.,
Banet,C.K., Frenkiel,M.P., Courageot,M.P., Coulibaly,F.,
Catteau,A., Flamand,M., Weber,P. and Ceccaldi,P.E.
TITLE Neurovirulent strain of the west nile virus and applications
thereof
JOURNAL Patent: WO 02081511-A 1 17-OCT-2002;
INSTITUT PASTEUR (FR) ; Kimron Veterinary Institute (IL)
FEATURES Location/Qualifiers
source 1. .11029
/organism="Flavivirus sp."
/mol_type="unassigned DNA"
/db_xref="taxon:11093"
CDS 97. .10398
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD60131.1"
/db_xref="GI:27646163"
/translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI

RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQVG
ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
ISRQELRCGNGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
QMWEAVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
KIIGTAVKNNLAISKLDSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPRLRCLNLDVYRIL
LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
ESDAEITGSSERVDVRLDDGENFQLMNDPGAPWKIWMLRMVCLAISAYTPWAILPSVV
GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
TEIPDRAWNSGYEWITEYTGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMPDNINMPNGLIAQFYQ
PEREKVYTMEGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
VFFLLMQRKGIGKIGLGGAVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHSVSRGTAKLRWLVER
RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKINNRIERLRREYSSTWHHDENH
PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
EKVDTKAPEPPEGAKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
EQNQWRSAREAVEDPKFWEMVDEERFAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
DEKYVDYMSSLKRYEDTTLVEDTVL"

ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 11029;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCGGGCTGTCAATATGCTAAA 21
||||||||||||||||||||
Db 129 CCGGGCTGTCAATATGCTAAA 149

RESULT 17
AX577796
LOCUS AX577796 11029 bp DNA linear PAT 08-JAN-2003

```
DEFINITION  Sequence 1 from Patent WO02081741.
ACCESSION   AX577796
VERSION     AX577796.1  GI:27647035
KEYWORDS    .
SOURCE      Flavivirus sp.
  ORGANISM  Flavivirus sp.
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus.
REFERENCE   1
  AUTHORS   Guenet,J.L., Mashimo,T., Simon-Chazottes,D., Montagutelli,X.,
            Frenkiel,M.P., Despres,P., Deubel,V., Bonhomme,F. and Lucas,M.
  TITLE     Use of products of genes of the 2'-5' oligoadenylate synthetase
            family (oas) for screening antiviral agents and for detecting
            responsiveness to flaviviridae infection
  JOURNAL   Patent: WO 02081741-A 1 17-OCT-2002;
            INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE
            SCIENTIFIQUE (CNRS) (FR)
FEATURES             Location/Qualifiers
     source          1. .11029
                     /organism="Flavivirus sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:11093"
     CDS             97. .10398
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD60138.1"
                     /db_xref="GI:27647036"
                     /translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
                     RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
                     KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
                     AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
                     VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
                     VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
                     MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
                     NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQVG
                     ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
                     MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
                     NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
                     PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
                     INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
                     RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
                     ISRQELRCGNGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
                     QMWEAVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
                     FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
                     KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
                     DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
                     TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
                     DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
                     GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
                     DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPRLRCLNLDVYRIL
                     LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
                     TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
                     ESDAEITGSSERVDVRLDDGENFQLMNDPGAPWKIWMLRMVCLAISAYTPWAILPSVV
                     GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
                     TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
                     VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
                     AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
                     LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
                     NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
                     TEIPDRAWNSGYEWITEYTGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
                     PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
                     SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMPDNINMPNGLIAQFYQ
                     PEREKVYTMEGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
                     TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
                     LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
                     VFFLLMQRKGIGKIGLGGAVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
                     QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
                     LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
                     GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
                     VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
                     TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
                     VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHSVSRGTAKLRWLVER
                     RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
                     KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
                     CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
                     RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKINNRIERLRREYSSTWHHDENH
                     PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
                     EKVDTKAPEPPEGAKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
                     EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
                     AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
                     YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
                     TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
                     RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
                     DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
                     QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
                     IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
                     DEKYVDYMSSLKRYEDTTLVEDTVL"
ORIGIN

  Query Match           100.0%;  Score 21;  DB 6;  Length 11029;
  Best Local Similarity  100.0%;  Pred. No. 1.6;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCTAAA 21
              |||||||||||||||||||||
Db        129 CCGGGCTGTCAATATGCTAAA 149


RESULT 18
AB185914
LOCUS       AB185914               11029 bp    RNA     linear   VRL 31-JUL-2004
DEFINITION  West Nile virus gene for polyprotein precursor protein, complete
            cds, isolate: 6-LP.
ACCESSION   AB185914
VERSION     AB185914.2  GI:50872124
KEYWORDS    .
SOURCE      West Nile virus (WNV)
  ORGANISM  West Nile virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE   1
  AUTHORS   Shirato,K., Miyoshi,H., Goto,A., Ako,Y., Ueki,T., Kariwa,H. and
            Takashima,I.
  TITLE     Correlation between viral envelope glycosylation and
            neuroinvasiveness of the New York strain of the West Nile virus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 11029)
  AUTHORS   Shirato,K., Kariwa,H. and Takashima,I.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUL-2004) Kazuya Shirato, Graduate School of
            Veterinary Medicine, Hokkaido University, Laboratory of Public
            Health, Depertment of Environmental Veterinary Medicine; Kita-19
            Nishi-9, Kita-ku, Sapporo, Hokkaido 060-0818, Japan
            (E-mail:shirato@vetmed.hokudai.ac.jp, Tel:81-11-706-5213(ex.5213),
            Fax:81-11-706-5213)
COMMENT     On Jul 30, 2004 this sequence version replaced gi:50838778.
FEATURES             Location/Qualifiers
     source          1. .11029
                     /organism="West Nile virus"
                     /mol_type="genomic RNA"
                     /strain="NY99-6922"
                     /isolate="6-LP"
                     /db_xref="taxon:11082"
     CDS             97. .10398
                     /codon_start=1
                     /product="polyprotein precursor protein"
                     /protein_id="BAD34488.1"
                     /db_xref="GI:50838779"
                     /translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
                     RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
                     KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
                     AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
                     VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
                     VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
                     MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
                     NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQVG
```

ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
QMWEAVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMVCLAISAYTPWAILPSVV
GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
TEIPDRAWNSGYEWITEYTGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
VFFLLMQRKGIGKIGLGGAVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
DEKYVDYMSSLKRYEDTTLVEDTVL"

ORIGIN

Query Match 100.0%; Score 21; DB 14; Length 11029;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCGGGCTGTCAATATGCTAAA 21
||||||||||||||||||||
Db 129 CCGGGCTGTCAATATGCTAAA 149

RESULT 19
AB185915
LOCUS AB185915 11029 bp RNA linear VRL 31-JUL-2004
DEFINITION West Nile virus gene for polyprotein precursor protein, complete cds, isolate: 6-SP.
ACCESSION AB185915
VERSION AB185915.2 GI:50872125
KEYWORDS .
SOURCE West Nile virus (WNV)
ORGANISM West Nile virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Flavivirus; Japanese encephalitis virus group.
REFERENCE 1
AUTHORS Shirato,K., Miyoshi,H., Goto,A., Ako,Y., Ueki,T., Kariwa,H. and Takashima,I.
TITLE Correlation between viral envelope glycosylation and neuroinvasiveness of the New York strain of the West Nile virus
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 11029)
AUTHORS Shirato,K., Kariwa,H. and Takashima,I.
TITLE Direct Submission
JOURNAL Submitted (28-JUL-2004) Kazuya Shirato, Graduate School of Veterinary Medicine, Hokkaido University, Laboratory of Public Health, Depertment of Environmental Veterinary Medicine; Kita-19 Nishi-9, Kita-ku, Sapporo, Hokkaido 060-0818, Japan (E-mail:shirato@vetmed.hokudai.ac.jp, Tel:81-11-706-5213(ex.5213), Fax:81-11-706-5213)
COMMENT On Jul 30, 2004 this sequence version replaced gi:50838780.
FEATURES Location/Qualifiers
source 1..11029
/organism="West Nile virus"
/mol_type="genomic RNA"
/strain="NY99-6922"
/isolate="6-SP"
/db_xref="taxon:11082"
CDS 97..10398
/codon_start=1
/product="polyprotein precursor protein"
/protein_id="BAD34489.1"
/db_xref="GI:50838781"
/translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYPTQVG
ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
QMWEAVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMVCLAISAYTPWAILPSVV
GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
TEIPDRAWNSGYEWITEYTGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
VFFLLMQRKGIGKIGLGGAVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV

TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
DEKYVDYMSSLKRYEDTTLVEDTVL"

ORIGIN

Query Match 100.0%; Score 21; DB 14; Length 11029;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCGGGCTGTCAATATGCTAAA 21
||||||||||||||||||||
Db 129 CCGGGCTGTCAATATGCTAAA 149

RESULT 20
AB185916
LOCUS AB185916 11029 bp RNA linear VRL 30-JUL-2004
DEFINITION West Nile virus gene for polyprotein precursor protein, complete cds, isolate: B-SP.
ACCESSION AB185916
VERSION AB185916.1 GI:50838782
KEYWORDS .
SOURCE West Nile virus (WNV)
ORGANISM West Nile virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Flavivirus; Japanese encephalitis virus group.
REFERENCE 1
AUTHORS Shirato,K., Miyoshi,H., Goto,A., Ako,Y., Ueki,T., Kariwa,H. and Takashima,I.
TITLE Correlation between viral envelope glycosylation and neuroinvasiveness of the New York strain of the West Nile virus
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 11029)
AUTHORS Shirato,K., Kariwa,H. and Takashima,I.
TITLE Direct Submission
JOURNAL Submitted (28-JUL-2004) Kazuya Shirato, Graduate School of Veterinary Medicine, Hokkaido University, Laboratory of Public Health, Depertment of Environmental Veterinary Medicine; Kita-19 Nishi-9, Kita-ku, Sapporo, Hokkaido 060-0818, Japan (E-mail:shirato@vetmed.hokudai.ac.jp, Tel:81-11-706-5213(ex.5213), Fax:81-11-706-5213)
FEATURES Location/Qualifiers
source 1. .11029
/organism="West Nile virus"
/mol_type="genomic RNA"
/strain="BC787"
/isolate="B-SP"
/db_xref="taxon:11082"
CDS 97. .10398
/codon_start=1
/product="polyprotein precursor protein"
/protein_id="BAD34490.1"
/db_xref="GI:50838783"
/translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVIMTVNATDVTDVITIPTAAGKNLCIVR
AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK

MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYFTQVG
ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
QMWEAVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMVCLAISAYTPWAILPSVV
GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
NLFVMDETHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
TEIPDRAWNSGYEWITEYTGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVPYHDRRWCFDGPR
TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
VFFLLMQRKGIGKIGLGGAVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
VVDGIVATDVPELERATPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
EQNQWRSAREAVEDPKFWEMVDEERFAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
DEKYVDYMSSLKRYEDTTLVEDTVL"

ORIGIN

Query Match 100.0%; Score 21; DB 14; Length 11029;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCGGGCTGTCAATATGCTAAA 21
||||||||||||||||||||
Db 129 CCGGGCTGTCAATATGCTAAA 149

RESULT 21
AB185917
LOCUS AB185917 11029 bp RNA linear VRL 30-JUL-2004
DEFINITION West Nile virus gene for polyprotein precursor protein, complete cds, isolate: B-LP.
ACCESSION AB185917
VERSION AB185917.1 GI:50838784
KEYWORDS .

```
SOURCE      West Nile virus (WNV)
  ORGANISM  West Nile virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE   1
  AUTHORS   Shirato,K., Miyoshi,H., Goto,A., Ako,Y., Ueki,T., Kariwa,H. and
            Takashima,I.
  TITLE     Correlation between viral envelope glycosylation and
            neuroinvasiveness of the New York strain of the West Nile virus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 11029)
  AUTHORS   Shirato,K., Kariwa,H. and Takashima,I.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUL-2004) Kazuya Shirato, Graduate School of
            Veterinary Medicine, Hokkaido University, Laboratory of Public
            Health, Depertment of Environmental Veterinary Medicine; Kita-19
            Nishi-9, Kita-ku, Sapporo, Hokkaido 060-0818, Japan
            (E-mail:shirato@vetmed.hokudai.ac.jp, Tel:81-11-706-5213(ex.5213),
            Fax:81-11-706-5213)
FEATURES             Location/Qualifiers
     source          1. .11029
                     /organism="West Nile virus"
                     /mol_type="genomic RNA"
                     /strain="BC787"
                     /isolate="B-LP"
                     /db_xref="taxon:11082"
     CDS             97. .10398
                     /codon_start=1
                     /product="polyprotein precursor protein"
                     /protein_id="BAD34491.1"
                     /db_xref="GI:50838785"
                     /translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
                     RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
                     KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVIMTVNATDVTDVITIPTAAGKNLCIVR
                     AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
                     VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
                     VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
                     MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
                     NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQVG
                     ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
                     MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
                     NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
                     PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
                     INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
                     RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
                     ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
                     QMWEAVKDELNTLLKENGVDFSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
                     FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
                     KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
                     DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
                     TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
                     DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
                     GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
                     DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
                     LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
                     TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
                     ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMVCLAISAYTPWAILPSVV
                     GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
                     TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
                     VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
                     AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
                     LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
                     NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
                     TEIPDRAWNSGYEWITEYTGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
                     PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
                     SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
                     PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVPYHDRRWCFDGPR
                     TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
                     LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
                     VFFLLMQRKGIGKIGLGGAVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
                     QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
                     LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
                     GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
                     VVDGIVATDVPELERATPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
                     TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
                     VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
                     RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
                     KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
                     CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
                     RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
                     PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
                     EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
                     EQNQWRSAREAVEDPKFWEMVDEERE AHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
                     AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
                     YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
                     TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
                     RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
                     DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
                     QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
                     IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
                     DEKYVDYMSSLKRYEDTTLVEDTVL"
ORIGIN

  Query Match           100.0%;  Score 21;  DB 14;  Length 11029;
  Best Local Similarity  100.0%;  Pred. No. 1.6;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCTAAA 21
              |||||||||||||||||||||
Db        129 CCGGGCTGTCAATATGCTAAA 149


RESULT 22
AF196835
LOCUS       AF196835               11029 bp    RNA     linear   VRL 07-DEC-2000
DEFINITION  West Nile virus strain NY99-flamingo382-99, complete genome.
ACCESSION   AF196835
VERSION     AF196835.2  GI:11597239
KEYWORDS    .
SOURCE      West Nile virus
  ORGANISM  West Nile virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE   1  (bases 1 to 11029)
  AUTHORS   Lanciotti,R.S., Roehrig,J.T., Deubel,V., Smith,J., Parker,M.,
            Steele,K., Crise,B., Volpe,K.E., Crabtree,M.B., Scherret,J.H.,
            Hall,R.A., MacKenzie,J.S., Cropp,C.B., Panigrahy,B., Ostlund,E.,
            Schmitt,B., Malkinson,M., Banet,C., Weissman,J., Komar,N.,
            Savage,H.M., Stone,W., McNamara,T. and Gubler,D.J.
  TITLE     Origin of the West Nile virus responsible for an outbreak of
            encephalitis in the northeastern United States
  JOURNAL   Science 286 (5448), 2333-2337 (1999)
  MEDLINE   20070288
   PUBMED   10600742
REFERENCE   2  (bases 1 to 11029)
  AUTHORS   Lanciotti,R., Roehrig,J., Volpe,K. and Panigrahy,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-1999) Division of Vector-Borne Diseases, Centers
            for Disease Control and Prevention, Rampart Road, Fort Collins, CO
            80521, USA
REFERENCE   3  (bases 1 to 11029)
  AUTHORS   Lanciotti,R., Roehrig,J., Volpe,K. and Panigrahy,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-DEC-2000) Division of Vector-Borne Diseases, Centers
            for Disease Control and Prevention, Rampart Road, Fort Collins, CO
            80521, USA
  REMARK    Sequence update by submitter
COMMENT     On Dec 7, 2000 this sequence version replaced gi:6636174.
FEATURES             Location/Qualifiers
     source          1. .11029
                     /organism="West Nile virus"
                     /mol_type="genomic RNA"
                     /strain="NY99-flamingo382-99"
                     /db_xref="taxon:11082"
                     /country="USA: New York"
                     /note="1999; passage level 0 in chicken embryo"
```

```
CDS             97. .10398
                /codon_start=1
                /product="polyprotein precursor"
                /protein_id="AAF20092.2"
                /db_xref="GI:11597240"
                /translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
                RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
                KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
                AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
                VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
                VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
                MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
                NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQVG
                ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
                MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
                NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
                PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
                INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
                RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
                ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
                QMWEAVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
                FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
                KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
                DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
                TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
                DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
                GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
                DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
                LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
                TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
                ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMVCLAISAYTPWAILPSVV
                GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
                TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
                VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
                AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
                LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
                NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
                TEIPDRAWNSGYEWITEYTGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
                PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
                SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
                PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
                TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
                LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
                VFFLLMQRKGIGKIGLGGAVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
                QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
                LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
                GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
                VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
                TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
                VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
                RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
                KSGVDVFYRPSECCDTLLCDIGESSSSAAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
                CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
                RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
                PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
                EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
                EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
                AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
                YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
                TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
                RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
                DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
                QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
                IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
                DEKYVDYMSSLKRYEDTTLVEDTVL"
mat_peptide     97. .465
                /product="nucleocapsid protein"
mat_peptide     466. .741
                /product="pre-membrane protein M"
mat_peptide     742. .966
                /product="membrane protein M"
mat_peptide     967. .2469
                /product="envelope glycoprotein E"
mat_peptide     2470. .3525
                /product="non-structural protein 1"
mat_peptide     3526. .4218
                /product="non-structural protein 2A"
mat_peptide     4219. .4611
                /product="non-structural protein 2B"
mat_peptide     4612. .6468
                /product="non-structural protein 3"
mat_peptide     6469. .6915
                /product="non-structural protein 4A"
mat_peptide     6916. .7680
                /product="non-structural protein 4B"
mat_peptide     7681. .10395
                /product="non-structural protein 5"
ORIGIN

  Query Match           100.0%;  Score 21;  DB 14;  Length 11029;
  Best Local Similarity  100.0%;  Pred. No. 1.6;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCTAAA 21
              |||||||||||||||||||||
Db        129 CCGGGCTGTCAATATGCTAAA 149


RESULT 23
AF260967
LOCUS       AF260967               11029 bp    RNA     linear   VRL 27-AUG-2000
DEFINITION  West Nile virus strain NY99-eqhs, complete genome.
ACCESSION   AF260967
VERSION     AF260967.1  GI:9930133
KEYWORDS    .
SOURCE      West Nile virus
  ORGANISM  West Nile virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Flavivirus; Japanese encephalitis virus group.
REFERENCE   1  (bases 1 to 11029)
  AUTHORS   Bowen,M., Meyer,R.F., McKinney,N., Morrill,W. and Lanciotti,R.
  TITLE     Complete genomic sequence of West Nile virus equine isolate New
            York 1999
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 11029)
  AUTHORS   Bowen,M., Meyer,R.F., McKinney,N., Morrill,W. and Lanciotti,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2000) Arbovirus Diseases Branch, Centers for
            Disease Control & Prevention, Rampart Road, Fort Collins, CO 80521,
            USA
FEATURES             Location/Qualifiers
     source          1. .11029
                     /organism="West Nile virus"
                     /mol_type="genomic RNA"
                     /strain="NY99-eqhs"
                     /specific_host="horse"
                     /db_xref="taxon:11082"
                     /country="USA: New York"
                     /note="1999"
     CDS             97. .10398
                     /codon_start=1
                     /product="polyprotein precursor"
                     /protein_id="AAG02038.1"
                     /db_xref="GI:9930134"
                     /translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
                     RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
                     KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
                     AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
                     VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
                     VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
                     MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
                     NGCGLFGKGSIDTCAKFACSTKAIGRTILKENIKYEVAIFVHGPTTVESHGNYSTQVG
                     ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
                     MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
                     NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
                     PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
```

INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
QMWEAVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAAFFQMAYH
DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMVCLAISAYTPWAILPSVV
GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
TEIPDRAWNSGYEWITEYTGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVPYHDRRWCFDGPR
TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
VFFLLMQRKGIGKIGLGGAVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDISSLFGQRIEVKENFSMGEFL
LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
VVDGIVATDVPELERATPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
DEKYVDYMSSLKRYEDTTLVEDTVL"
mat_peptide 97. .465
/product="nucleocapsid protein C"
mat_peptide 466. .741
/product="pre-membrane protein prM"
mat_peptide 742. .966
/product="membrane protein M"
mat_peptide 967. .2469
/product="envelope glycoprotein E"
mat_peptide 2470. .3525
/product="non-structural protein 1 NS1"
mat_peptide 3526. .4218
/product="non-structural protein 2A NS2A"
mat_peptide 4219. .4611
/product="non-structural protein 2B NS2B"
mat_peptide 4612. .6468
/product="non-structural protein 3 NS3"
mat_peptide 6469. .6915
/product="non-structural protein 4A NS4A"
mat_peptide 6916. .7680
/product="non-structural protein NS4B"
mat_peptide 7681. .10395
/product="non-structural protein NS5"
ORIGIN

Query Match 100.0%; Score 21; DB 14; Length 11029;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCGGGCTGTCAATATGCTAAA 21
||||||||||||||||||||
Db 129 CCGGGCTGTCAATATGCTAAA 149

RESULT 24
AF260968
LOCUS AF260968 11029 bp RNA linear VRL 27-AUG-2000
DEFINITION West Nile virus strain Eg101, complete genome.
ACCESSION AF260968
VERSION AF260968.1 GI:9930135
KEYWORDS .
SOURCE West Nile virus
ORGANISM West Nile virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Flavivirus; Japanese encephalitis virus group.
REFERENCE 1 (bases 1 to 11029)
AUTHORS Bowen,M., Meyer,R.F., McKinney,N., Morrill,W. and Lanciotti,R.
TITLE Complete genomic sequence of West Nile virus strain Eg101
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 11029)
AUTHORS Bowen,M., Meyer,R.F., McKinney,N., Morrill,W. and Lanciotti,R.
TITLE Direct Submission
JOURNAL Submitted (27-APR-2000) Arbovirus Diseases Branch, Centers for
Disease Control & Prevention, Rampart Road, Fort Collins, CO 80521,
USA
FEATURES Location/Qualifiers
source 1. .11029
/organism="West Nile virus"
/mol_type="genomic RNA"
/strain="Eg101"
/db_xref="taxon:11082"
CDS 97. .10398
/codon_start=1
/product="polyprotein precursor"
/protein_id="AAG02039.1"
/db_xref="GI:9930136"
/translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCKQGVVDRGWG
NGCGLFGKGSIDTCAKFACSTKATGRTILKENIKYEVAIFVHGPTTVESHGNYPTQIG
ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
QMWEAVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
KIIGTAVKNNLAIHSDLSYWIESRLNDTWKLERAVLGEVKSCTWPETHTLWGEGILES
DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMACLAISAYTPWAILPSVV
GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
TEIPDRAWNSGYEWITEYIGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY

PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSLVMTMG
VFFLLMQRKGIGKIGLGGVVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
QRSQTDNQLAVFLICVLTLVSAVAANEMGWLDKTKNDISSLFGQRIEAKENFSMGEFL
LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLIKNMEKPGLKRGGAKGRTLGE
VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM
KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPKEFCVKVL
CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIKKVNSNAALGAMFE
EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRTRATWAENIQVAINQVRAIIG
DEKYVDYMSSLKRYEDTTLVEDTVL"

mat_peptide	97. .465	
	/product="nucleocapsid protein C"	
mat_peptide	466. .741	
	/product="pre-membrane protein prM"	
mat_peptide	742. .966	
	/product="membrane protein M"	
mat_peptide	967. .2469	
	/product="envelope glycoprotein E"	
mat_peptide	2470. .3525	
	/product="non-structural protein 1 NS1"	
mat_peptide	3526. .4218	
	/product="non-structural protein 2A NS2A"	
mat_peptide	4219. .4611	
	/product="non-structural protein 2B NS2B"	
mat_peptide	4612. .6468	
	/product="non-structural protein 3 NS3"	
mat_peptide	6469. .6915	
	/product="non-structural protein 4A NS4A"	
mat_peptide	6916. .7680	
	/product="non-structural protein NS4B"	
mat_peptide	7681. .10395	
	/product="non-structural protein NS5"	

ORIGIN

Query Match 100.0%; Score 21; DB 14; Length 11029;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CCGGGCTGTCAATATGCTAAA 21
              |||||||||||||||||||||
Db        129 CCGGGCTGTCAATATGCTAAA 149
```

RESULT 25
AF260969
LOCUS AF260969 11029 bp RNA linear VRL 27-AUG-2000
DEFINITION West Nile virus strain RO97-50, complete genome.
ACCESSION AF260969
VERSION AF260969.1 GI:9930137
KEYWORDS .
SOURCE West Nile virus
ORGANISM West Nile virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Flavivirus; Japanese encephalitis virus group.
REFERENCE 1 (bases 1 to 11029)
AUTHORS Savage,H.M., Ceianu,C., Nicolescu,G., Karabatsos,N., Lanciotti,R., Vladimirescu,A., Laiv,L., Ungureanu,A., Romanca,C. and Tsai,T.F.
TITLE Entomologic and avian investigations of an epidemic of West Nile fever in Romania in 1996, with serologic and molecular characterization of a virus isolate from mosquitoes
JOURNAL Am. J. Trop. Med. Hyg. 61 (4), 600-611 (1999)
MEDLINE 20014331
PUBMED 10548295
REFERENCE 2 (bases 1 to 11029)
AUTHORS Bowen,M., Meyer,R.F., McKinney,N., Morrill,W. and Lanciotti,R.
TITLE Direct Submission
JOURNAL Submitted (27-APR-2000) Arbovirus Diseases Branch, Centers for Disease Control & Prevention, Rampart Road, Fort Collins, CO 80521, USA

FEATURES Location/Qualifiers

source	1. .11029	
	/organism="West Nile virus"	
	/mol_type="genomic RNA"	
	/strain="RO97-50"	
	/specific_host="Culex pipiens"	
	/db_xref="taxon:11082"	
	/country="Romania"	
	/note="1996"	
CDS	97. .10398	
	/codon_start=1	
	/product="polyprotein precursor"	
	/protein_id="AAG02040.1"	
	/db_xref="GI:9930138"	

/translation="MSKKPGGPGKSRAVNMLKRGMPRVLSLIGLKRAMLSLIDGKGPI
RFVLALLAFFRFTAIAPTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSSKQ
KKRGGKTGIAVMIGLIASVGAVTLSNFQGKVMMTVNATDVTDVITIPTAAGKNLCIVR
AMDVGYMCDDTITYECPVLSAGNDPEDIDCWCTKSAVYVRYGRCTKTRHSRRSRRSLT
VQTHGESTLANKKGAWMDSTKATRYLVKTESWILRNPGYALVAAVIGWMLGSNTMQRV
VFVVLLLLVAPAYSFNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVK
MMNMEAANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRGWG
NGCGLFGKGSIDTCAKFACSTKATGRTILKENIKYEVAIFVHGPTTVESHGNYSTQIG
ATQAGRFSITPAAPSYTLKLGEYGEVTVDCEPRSGIDTNAYYVMTVGTKTFLVHREWF
MDLNLPWSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSS
NTVKLTSGHLKCRVKMEKLQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDG
PCKVPISSVASLNDLTPVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQ
INHHWHKSGSSIGKAFTTTLKGAQRLAALGDTAWDFGSVGGVFTSVGKAVHQVFGGAF
RSLFGGMSWITQGLLGALLLWMGINARDRSIALTFLAVGGVLLFLSVNVHADTGCAID
ISRQELRCGSGVFIHNDVEAWMDRYKYYPETPQGLAKIIQKAHKEGVCGLRSVSRLEH
QMWESVKDELNTLLKENGVDLSVVVEKQEGMYKSAPKRLTATTEKLEIGWKAWGKSIL
FAPELANNTFVVDGPETKECPTQNRAWNSLEVEDFGFGLTSTRMFLKVRESNTTECDS
KIIGTAVKNNLAIHSDLSYWIESRFNDTWKLERAVLGEVKSCTWPETHTLWGDGILES
DLIIPVTLAGPRSNHNRRPGYKTQNQGPWDEGRVEIDFDYCPGTTVTLSESCGHRGPA
TRTTTESGKLITDWCCRSCTLPPLRYQTDSGCWYGMEIRPQRHDEKTLVQSQVNAYNA
DMIDPFQLGLLVVFLATQEVLRKRWTAKISMPAILIALLVLVFGGITYTDVLRYVILV
GAAFAESNSGGDVVHLALMATFKIQPVFMVASFLKARWTNQENILLMLAAVFFQMAYH
DARQILLWEIPDVLNSLAVAWMILRAITFTTTSNVVVPLLALLTPGLRCLNLDVYRIL
LLMVGIGSLIREKRSAAAKKKGASLLCLALASTGLFNPMILAAGLIACDPNRKRGWPA
TEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISGKSTDMWIERTADISW
ESDAEITGSSERVDVRLDDDGNFQLMNDPGAPWKIWMLRMVCLAVSAYTPWAILPSVV
GFWITLQYTKRGGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGQDEVQMIVVEPGKN
VKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYIS
AIVQGERMDEPIPAGFEPEMLRKKQITVLDLHPGAGKTRRILPQIIKEAINRRLRTAV
LAPTRVVAAEMAEALRGLPIRYQTSAVPREHNGNEIVDVMCHATLTHRLMSPHRVPNY
NLFVMDEAHFTDPASIAARGYISTKVELGEAAAIFMTATPPGTSDPFPESNSPISDLQ
TEIPDRAWNSGYEWITEYIGKTVWFVPSVKMGNEIALCLQRAGKKVVQLNRKSYETEY
PKCKNDDWDFVITTDISEMGANFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAA
SAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLIAQFYQ
PEREKVYTMDGEYRLRGEERKNFLELLRTADLPVWLAYKVAAAGVSYHDRRWCFDGPR
TNTILEDNNEVEVITKLGERKILRPRWIDARVYSDHQALKAFKDFASGKRSQIGLIEV
LGKMPEHFMGKTWEALDTMYVVATAEKGGRAHRMALEELPDALQTIALIALLSVMTMG
VFFLLMQRKGIGKIGLGGVVLGVATFFCWMAEVPGTKIAGMLLLSLLLMIVLIPEPEK
QRSQTDNQLAVFLICVMTLVSAVAANEMGWLDKTKSDVSSLFGQRIEVKENFSMGEFL
LDLRPATAWSLYAVTTAVLTPLLKHLITSDYINTSLTSINVQASALFTLARGFPFVDV
GVSALLLAAGCWGQVTLTVTVTAATLLFCHYAYMVPGWQAEAMRSAQRRTAAGIMKNA
VVDGIVATDVPELERTTPIMQKKVGQIMLILVSLAAVVVNPSVKTVREAGILITAAAV
TLWENGASSVWNATTAIGLCHIMRGGWLSCLSITWTLMKNMDKPGLKRGGAKGRTLGE
VWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVER
RFLEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTM

```
                     KSGVDVFYRPSECCDTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVL
                     CPYMPKVIEKMELLQRRYGGGLVRNPLSRNSTHEMYWVSRASGNVVHSVNMTSQVLLG
                     RMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIKNRIERLRREYSSTWHHDENH
                     PYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTPFGQQRVFK
                     EKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFE
                     EQNQWRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSR
                     AIWFMWLGARFLEFEALGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGTRPGGKI
                     YADDTAGWDTRITRADLENEAKVLELLDGEHRRLARAIIELTYRHKVVKVMRPAADGR
                     TVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEGVIGPDDVEKLTKGKGPKV
                     RTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWKPSTGWY
                     DWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYA
                     QMWLLLYFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVW
                     IEENEWMEDKTPVEKWSDVPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIG
                     DEKYVDYMSSLKRYEDTILVEDTVL"
     mat_peptide     97. .465
                     /product="nucleocapsid protein C"
     mat_peptide     466. .741
                     /product="pre-membrane protein prM"
     mat_peptide     742. .966
                     /product="membrane protein M"
     mat_peptide     967. .2469
                     /product="envelope glycoprotein E"
     mat_peptide     2470. .3525
                     /product="non-structural protein 1 NS1"
     mat_peptide     3526. .4218
                     /product="non-structural protein 2A NS2A"
     mat_peptide     4219. .4611
                     /product="non-structural protein 2B NS2B"
     mat_peptide     4612. .6468
                     /product="non-structural protein 3 NS3"
     mat_peptide     6469. .6915
                     /product="non-structural protein 4A NS4A"
     mat_peptide     6916. .7680
                     /product="non-structural protein NS4B"
     mat_peptide     7681. .10395
                     /product="non-structural protein NS5"
ORIGIN

  Query Match           100.0%;  Score 21;  DB 14;  Length 11029;
  Best Local Similarity  100.0%;  Pred. No. 1.6;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCTAAA 21
              |||||||||||||||||||||
Db        129 CCGGGCTGTCAATATGCTAAA 149


Search completed: September  6, 2005, 20:29:40
Job time : 740.656 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 6, 2005, 16:01:23 ; Search time 189.656 Seconds
(without alignments)
655.473 Million cell updates/sec

Title: US-10-729-421-34
Perfect score: 21
Sequence: 1 ccgggctgtcaatatgctaaa 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	21	100.0	21	12	ADQ30664	Adq30664 West Nile
	2	21	100.0	37	12	ADN36773	Adn36773 West Nile
	3	21	100.0	65	12	ADN36771	Adn36771 West Nile
	4	21	100.0	365	6	ABK51710	Abk51710 Partial c
	5	21	100.0	366	8	ABQ76684	Abq76684 WNVCwt DN
	6	21	100.0	967	12	ADQ30647	Adq30647 West Nile
	7	21	100.0	10945	13	ADR32078	Adr32078 Genomic D
	8	21	100.0	10945	13	ADR67768	Adr67768 West Nile
	9	21	100.0	10962	12	ADK13681	Adk13681 West Nile
	10	21	100.0	10975	12	ADN98022	Adn98022 West Nile
	11	21	100.0	11029	8	ABZ68481	Abz68481 Nucleotid
	12	21	100.0	11029	10	ABV74821	Abv74821 West Nile
	13	21	100.0	11029	12	ADN98023	Adn98023 West Nile
c	14	19	90.5	21	12	ADN36695	Adn36695 West Nile
c	15	19	90.5	24	12	ADN36823	Adn36823 West Nile
c	16	19	90.5	48	12	ADN36707	Adn36707 West Nile
c	17	19	90.5	69	12	ADN36694	Adn36694 West Nile
c	18	18	85.7	20	12	ADN36696	Adn36696 West Nile
c	19	18	85.7	47	12	ADN36708	Adn36708 West Nile
	20	17.8	84.8	4512	2	AAQ22767	Aaq22767 JEV Nakay
	21	17.8	84.8	10818	12	ADO07431	Ado07431 Japanese
	22	17.8	84.8	10968	12	ADO07437	Ado07437 Japanese
	23	17.8	84.8	10976	3	ABL50890	Abl50890 Japanese
	24	17.8	84.8	18563	12	ADO07466	Ado07466 Japanese
	25	17.8	84.8	18563	12	ADO07465	Ado07465 Japanese
	26	17.8	84.8	18565	12	ADO07467	Ado07467 Japanese
	27	17.8	84.8	19038	12	ADO07468	Ado07468 Japanese
	28	17.8	84.8	19038	12	ADO07469	Ado07469 Japanese
	29	17.8	84.8	19040	12	ADO07470	Ado07470 Japanese
	30	17	81.0	17	6	ACN01436	Acn01436 WNV Inozy
	31	17	81.0	17	6	ACN00029	Acn00029 WNV Hamme
c	32	17	81.0	17	6	ACN09480	Acn09480 WNV minus
c	33	17	81.0	17	6	ACN15271	Acn15271 WNV minus
c	34	17	81.0	17	6	ACN14165	Acn14165 WNV minus
	35	17	81.0	17	6	ACN04724	Acn04724 WNV DNAzy
c	36	17	81.0	17	6	ACN09479	Acn09479 WNV minus
	37	17	81.0	20	12	ADN36776	Adn36776 West Nile
	38	16.8	80.0	2262	8	ACA54320	Aca54320 Prokaryot
c	39	16.4	78.1	356	9	ACH31834	Ach31834 Human bon
	40	16.2	77.1	326	6	ABN75146	Abn75146 Human ORF
	41	16.2	77.1	1149	10	ADB46059	Adb46059 rapC DNA
	42	16.2	77.1	2023	10	ADB69062	Adb69062 C. neofor
c	43	16.2	77.1	2232	11	ACH98107	Ach98107 Klebsiell
	44	16.2	77.1	3177	4	ABL16411	Abl16411 Drosophil
c	45	16.2	77.1	7018	4	ABL16410	Abl16410 Drosophil
	46	16	76.2	17	6	ACN00030	Acn00030 WNV Hamme
	47	16	76.2	17	6	ACN03477	Acn03477 WNV Zinzy
c	48	16	76.2	2005	4	AAF81805	Aaf81805 Human sec
	49	16	76.2	149671	6	ABK84797	Abk84797 Human cDN
	50	16	76.2	149671	9	ADB70361	Adb70361 Moesin cD
	51	16	76.2	149671	12	ADJ37140	Adj37140 Human mal
	52	15.8	75.2	624	6	ABN73108	Abn73108 Bovine em
	53	15.8	75.2	1197	5	AAS81962	Aas81962 DNA encod
c	54	15.8	75.2	1208	5	AAS74882	Aas74882 DNA encod
	55	15.8	75.2	1208	5	AAS93301	Aas93301 DNA encod
	56	15.8	75.2	1208	5	AAS77346	Aas77346 DNA encod
	57	15.8	75.2	2373	5	AAS86894	Aas86894 DNA encod
c	58	15.8	75.2	11184	12	ADP86274	Adp86274 Hepatitis
c	59	15.8	75.2	11184	12	ADP86276	Adp86276 Hepatitis
c	60	15.8	75.2	11184	12	ADP86277	Adp86277 Hepatitis
c	61	15.8	75.2	11313	12	ADP86273	Adp86273 Hepatitis
c	62	15.8	75.2	11313	12	ADP86264	Adp86264 Hepatitis
c	63	15.8	75.2	11313	12	ADP86266	Adp86266 Hepatitis
c	64	15.8	75.2	11313	12	ADP86265	Adp86265 Hepatitis
c	65	15.8	75.2	11313	12	ADP86268	Adp86268 Hepatitis
c	66	15.8	75.2	11313	12	ADP86270	Adp86270 Hepatitis
c	67	15.8	75.2	11313	12	ADP86271	Adp86271 Hepatitis
c	68	15.8	75.2	11313	12	ADP86272	Adp86272 Hepatitis
c	69	15.8	75.2	11313	12	ADP86269	Adp86269 Hepatitis
c	70	15.8	75.2	11313	12	ADP86275	Adp86275 Hepatitis
c	71	15.8	75.2	11313	12	ADP86267	Adp86267 Hepatitis
c	72	15.8	75.2	12306	10	ADI41414	Adi41414 BB7 nucle
c	73	15.8	75.2	12315	10	ADI41413	Adi41413 BB7M4hRLu
c	74	15.8	75.2	12980	2	AAV59364	Aav59364 Hepatitis
c	75	15.8	75.2	12980	6	ABK87286	Abk87286 Hepatitis
c	76	15.8	75.2	12980	8	ACA62469	Aca62469 DNA encod
c	77	15.8	75.2	15065	3	AAZ36195	Aaz36195 Nucleotid
	78	15.8	75.2	16847	12	ADO07464	Ado07464 Japanese
c	79	15.4	73.3	444	9	ACH24983	Ach24983 Human adu
	80	15.4	73.3	1119	6	AAD31757	Aad31757 Soybean H
c	81	15.4	73.3	73583	12	ADQ59187	Adq59187 MSI-H car
c	82	15.4	73.3	222930	6	ABK84349	Abk84349 Human cDN
	83	15.4	73.3	295096	11	ACN44068	Acn44068 Mouse gen
	84	15.2	72.4	26	12	ADN36839	Adn36839 West Nile
	85	15.2	72.4	445	4	AAS36425	Aas36425 Human car
	86	15.2	72.4	445	10	ADE47119	Ade47119 Human car
	87	15.2	72.4	445	13	ADJ08537	Adj08537 Human car
c	88	15.2	72.4	452	3	AAA82296	Aaa82296 N. mening
	89	15.2	72.4	914	5	AAS66345	Aas66345 DNA encod
c	90	15.2	72.4	1038	3	AAZ60397	Aaz60397 A diacylg
	91	15.2	72.4	1065	12	ADP72956	Adp72956 Renal tox
	92	15.2	72.4	1121	2	AAQ21554	Aaq21554 Polyfunct
	93	15.2	72.4	1121	6	ABK63744	Abk63744 Rat seque

94	15.2	72.4	1121	10	ADB58300	Adb58300 Toxicity-
95	15.2	72.4	1121	10	ADB52851	Adb52851 Primary r
96	15.2	72.4	1121	10	ABT41990	Abt41990 Toxicity
97	15.2	72.4	1167	10	ADF01035	Adf01035 Bacterial
c 98	15.2	72.4	1302	4	AAF58403	Aaf58403 Rat rOCIL
c 99	15.2	72.4	1389	11	ACH96542	Ach96542 Klebsiell
100	15.2	72.4	1418	6	ABZ15428	Abz15428 Arabidops

ALIGNMENTS

RESULT 1
ADQ30664
ID ADQ30664 standard; DNA; 21 BP.
XX
AC ADQ30664;
XX
DT 23-SEP-2004 (first entry)
XX
DE West Nile Virus capsid gene sense primer WNVVA1.
XX
KW ss; primer; West Nile Virus; diagnosis.
XX
OS West Nile virus.
XX
PN WO2004055159-A2.
XX
PD 01-JUL-2004.
XX
PF 05-DEC-2003; 2003WO-US038750.
XX
PR 12-DEC-2002; 2002US-0432850P.
PR 20-JUN-2003; 2003US-0480431P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Shyamala V;
XX
DR WPI; 2004-488058/46.
XX
PT New isolated oligonucleotides for accurately diagnosing West Nile virus
PT infection or for capturing, detecting and quantitating West Nile virus in
PT blood samples.
XX
PS Claim 1; SEQ ID NO 34; 56pp; English.
XX
CC The invention relates to an isolated oligonucleotide not more than 60
CC nucleotides in length comprising a nucleotide sequence (S1) of at least
CC 10 contiguous nucleotides from any of the 28 nucleotide sequences (e.g.
CC 20, 21 or 23 bp) given in the specification derived from the West Nile
CC Virus (WNV) genome, a nucleotide sequence (S2) having 90% sequence
CC identity to the nucleotide sequence of (S1), or complements of (S1) and
CC (S2). The oligonucleotide further comprises a detectable label at the 5'-
CC end and/or the 3'-end. The detectable label is a fluorescent label
CC selected from 6-carboxyfluorescein (6-FAM), tetramethyl rhodamine
CC (TAMRA), and 2',4',5',7',-tetrachloro-4-7-dichlorofluorescein (TET). The
CC composition and methods are useful for accurately diagnosing West Nile
CC virus infection or for capturing, detecting and quantitating West Nile
CC virus in biological samples, particularly blood samples. This sequence
CC corresponds to a PCR primer to amplify a fragment of the capsid gene of
CC the WNV genome. The fragment is detected using the oligonucleotides of
CC the invention.
XX
SQ Sequence 21 BP; 6 A; 5 C; 5 G; 5 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 12; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.38;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCGGGCTGTCAATATGCTAAA 21
||||||||||||||||||||
Db 1 CCGGGCTGTCAATATGCTAAA 21

RESULT 2
ADN36773
ID ADN36773 standard; DNA; 37 BP.
XX
AC ADN36773;
XX
DT 15-JUL-2004 (first entry)
XX
DE West Nile virus detection-related oligonucleotide probe SeqID95.
XX
KW hybridisation assay probe; nucleic acid detection;
KW target-complementary sequence; flavivirus; West Nile virus; WNV;
KW RNA virus; infection; meningitis; encephalitis;
KW high throughput screening; probe; ss.
XX
OS West Nile virus.
XX
PN WO2004036190-A2.
XX
PD 29-APR-2004.
XX
PF 10-OCT-2003; 2003WO-US033639.
XX
PR 16-OCT-2002; 2002US-0418891P.
PR 25-NOV-2002; 2002US-0429006P.
PR 24-FEB-2003; 2003US-0449810P.
XX
PA (GENP-) GEN-PROBE INC.
XX
PI Linnen JM, Pollner RB, Wu W, Dennis GG, Darby PM;
XX
DR WPI; 2004-389590/36.
XX
PT New hybridization assay probe comprising target-complementary sequence of
PT bases, useful in detecting flavivirus, e.g. West Nile virus.
XX
PS Claim 55; SEQ ID NO 95; 135pp; English.
XX
CC This invention relates to a novel hybridisation assay probe, for
CC detecting a nucleic acid, which is a probe sequence that comprises a
CC target-complementary sequence of bases, and optionally one or more base
CC sequences that are not complementary to the nucleic acid that is to be
CC detected. The hybridisation assay probes and the kits are useful in
CC detecting and amplifying a target nucleic acid sequence, for example
CC flavivirus like West Nile virus, that may be present in a biological
CC sample. West Nile virus (WNV) is an RNA virus that primarily infects
CC birds and culex mosquitos, with humans and horses serving as incidental
CC hosts. Infection of humans can lead to meningitis or encephalitis. The
CC invention may allow for accurate and efficient high throughput screening.
CC The present sequence is that of an oligonucleotide probe which is related
CC to the invention.
XX
SQ Sequence 37 BP; 10 A; 10 C; 11 G; 6 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 12; Length 37;
Best Local Similarity 100.0%; Pred. No. 0.41;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCGGGCTGTCAATATGCTAAA 21
||||||||||||||||||||
Db 4 CCGGGCTGTCAATATGCTAAA 24

RESULT 3
ADN36771
ID ADN36771 standard; DNA; 65 BP.
XX
AC ADN36771;
XX
DT 15-JUL-2004 (first entry)
XX

DE West Nile virus detection-related oligonucleotide probe SeqID93.
XX
KW hybridisation assay probe; nucleic acid detection;
KW target-complementary sequence; flavivirus; West Nile virus; WNV;
KW RNA virus; infection; meningitis; encephalitis;
KW high throughput screening; probe; ss.
XX
OS West Nile virus.
XX
PN WO2004036190-A2.
XX
PD 29-APR-2004.
XX
PF 10-OCT-2003; 2003WO-US033639.
XX
PR 16-OCT-2002; 2002US-0418891P.
PR 25-NOV-2002; 2002US-0429006P.
PR 24-FEB-2003; 2003US-0449810P.
XX
PA (GENP-) GEN-PROBE INC.
XX
PI Linnen JM, Pollner RB, Wu W, Dennis GG, Darby PM;
XX
DR WPI; 2004-389590/36.
XX
PT New hybridization assay probe comprising target-complementary sequence of
PT bases, useful in detecting flavivirus, e.g. West Nile virus.
XX
PS Example 6; SEQ ID NO 93; 135pp; English.
XX
CC This invention relates to a novel hybridisation assay probe, for
CC detecting a nucleic acid, which is a probe sequence that comprises a
CC target-complementary sequence of bases, and optionally one or more base
CC sequences that are not complementary to the nucleic acid that is to be
CC detected. The hybridisation assay probes and the kits are useful in
CC detecting and amplifying a target nucleic acid sequence, for example
CC flavivirus like West Nile virus, that may be present in a biological
CC sample. West Nile virus (WNV) is an RNA virus that primarily infects
CC birds and culex mosquitos, with humans and horses serving as incidental
CC hosts. Infection of humans can lead to meningitis or encephalitis. The
CC invention may allow for accurate and efficient high throughput screening.
CC The present sequence is that of an oligonucleotide probe which is related
CC to the invention.
XX
SQ Sequence 65 BP; 19 A; 17 C; 20 G; 9 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 12; Length 65;
Best Local Similarity 100.0%; Pred. No. 0.44;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCGGGCTGTCAATATGCTAAA 21
||||||||||||||||||||
Db 32 CCGGGCTGTCAATATGCTAAA 52

RESULT 4
ABK51710
ID ABK51710 standard; cDNA; 365 BP.
XX
AC ABK51710;
XX
DT 27-AUG-2002 (first entry)
XX
DE Partial cDNA for west nile virus capsid protein.
XX
KW Human; ss; IgE leader sequence; west nile virus capsid protein;
KW RNA secondary structure; free energy; gene therapy; cancer;
KW hyperproliferative disease; autoimmune disease; rheumatoid arthritis;
KW multiple sclerosis; Sjogren's syndrome; sarcoidosis; scleroderma;
KW insulin-dependent diabetes mellitus; autoimmune thyroiditis; psoriasis;
KW reactive arthritis; ankylosing spondylitis; polymyositis; vasculitis;
KW dermatomyositis; Crohn's disease; ulcerative colitis.
XX
OS West Nile virus.
XX
PN WO200229088-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-US031451.
XX
PR 04-OCT-2000; 2000US-0237885P.
XX
PA (UYPE-) UNIV PENNSYLVANIA.
XX
PI Weiner DB, Yang J;
XX
DR WPI; 2002-416682/44.
XX
PT Producing recombinant protein for preparing pharmaceutical compounds to
PT treat, e.g., cancers or autoimmune disorders, comprises predicting
PT secondary structure (SS) of mRNA and modifying DNA to give mRNA with SS
PT having increased free energy.
XX
PS Example 2; Fig 1; 48pp; English.
XX
CC The invention relates to producing (M1) a protein (I) in a recombinant
CC expression system (II) comprising: (a) predicting the secondary structure
CC of mRNA; (b) modifying the native heterologous DNA sequence where the
CC mRNA transcribed from the modified DNA has a secondary structure with
CC increased free energy; and (c) using the modified DNA in (II) for
CC production of (I). Also included are (1) an injectable pharmaceutical
CC composition comprising a nucleic acid molecule that includes a modified
CC coding sequence (IV) encoding a protein operably linked to regulatory
CC elements, where (IV) comprises a higher AT or AU content relative to the
CC AT or AU content of the native coding sequence and further comprising a
CC pharmaceutical carrier and (2) a recombinant viral vector comprising a
CC nucleic acid molecule that includes (IV). The method is used for
CC producing a protein in a recombinant expression system. Use of a nucleic
CC acid or recombinant viral vector that have modified DNA sequences to
CC improve protein production can be used in gene therapy and for the
CC treatment of cancers, hyperproliferative diseases, and autoimmune
CC diseases such as rheumatoid arthritis, multiple sclerosis, Sjogren's
CC syndrome, sarcoidosis, insulin-dependent diabetes mellitus, autoimmune
CC thyroiditis, reactive arthritis, ankylosing spondylitis, scleroderma,
CC polymyositis, dermatomyositis, psoriasis, vasculitis, Crohn's disease and
CC ulcerative colitis. The present sequence is a cDNA for West Nile virus
CC capsid protein. Fusion constructs of modified mRNA for the capsid protein
CC and human IgE leader sequence are used in an experiment to minimise the
CC free energy of the capsid protein mRNA. Note:The present sequence is not
CC shown in the specification but was created using the information in
CC figure 1 and the sequence appearing as ABK51708
XX
SQ Sequence 365 BP; 103 A; 80 C; 109 G; 73 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 6; Length 365;
Best Local Similarity 100.0%; Pred. No. 0.57;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCGGGCTGTCAATATGCTAAA 21
||||||||||||||||||||
Db 29 CCGGGCTGTCAATATGCTAAA 49

RESULT 5
ABQ76684
ID ABQ76684 standard; DNA; 366 BP.
XX
AC ABQ76684;
XX
DT 13-MAY-2003 (first entry)
XX
DE WNVCwt DNA fragment.
XX
KW Capsid protein; WNVCwt; mRNA secondary structure; cancer;
KW immunosuppressive; antirheumatic; cytostatic; antiulcer; neuroprotective;

```
KW   antiarthritic; antidiabetic; antithyroid; antipsoriatic; virucide; gene;
KW   antiparasitic; antiallergic; gene therapy; allergen; multiple sclerosis;
KW   protective immune response; hyperproliferative cell; ulcerative colitis;
KW   hyperproliferative disease; psoriasis; autoimmune disease; psoriasis;
KW   rheumatoid arthritis; Sjogren's syndrome; autoimmune thyroiditis;
KW   insulin dependent diabetes mellitus; Crohn's disease; ds.
XX
OS   West nile virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .366
FT                   /*tag=  a
FT                   /product= "WNVCwt"
FT                   /note= "no start or stop codon given"
XX
PN   US2002123099-A1.
XX
PD   05-SEP-2002.
XX
PF   04-OCT-2001; 2001US-00971806.
XX
PR   04-OCT-2000; 2000US-0237885P.
XX
PA   (WEIN/) WEINER D B.
PA   (YANG/) YANG J.
XX
PI   Weiner DB,  Yang J;
XX
DR   WPI; 2003-066795/06.
DR   P-PSDB; ABG73556.
XX
PT   Producing protein in recombinant expression system involves predicting
PT   secondary structure of RNA encoding a protein and increasing free energy
PT   for the secondary structure by modifying sequence of DNA encoding the
PT   RNA.
XX
PS   Example 2; Fig 1; 25pp; English.
XX
CC   This invention describes a novel method for producing a protein by
CC   translation of mRNA from heterologous DNA sequences. The method involves
CC   predicting the secondary structure of mRNA transcribed from a native
CC   heterologous DNA sequence, modifying the sequence where mRNA transcribed
CC   from the modified DNA sequence has a secondary structure with increased
CC   free energy compared to mRNA transcribed from native DNA and using
CC   modified heterologous DNA for protein production. The products of the
CC   invention have immunosuppressive, antirheumatic, cytostatic, antiulcer,
CC   neuroprotective, antiarthritic, antidiabetic, antithyroid, antipsoriatic,
CC   virucide, antiparasitic and antiallergic activity and can be used for
CC   gene therapy. The method described is useful for producing a protein in a
CC   recombinant expression system, preferably a cell free in vitro
CC   transcription and translation system, an in vitro cell expression system,
CC   a DNA construct used in direct DNA injection, or a recombinant vector for
CC   delivery of DNA to an individual. The products of the invention are
CC   useful for eliciting broad immune responses against a target protein,
CC   i.e. proteins specifically associated with pathogens such as viruses,
CC   parasites, allergens, or the individual's own abnormal cells.
CC   Compositions containing the products of the invention confer a broad
CC   based protective immune response against hyperproliferative cells that
CC   are characteristic in hyperproliferative diseases including all forms of
CC   cancer and psoriasis. Such compositions are also useful for treating
CC   individuals suffering from autoimmune diseases including rheumatoid
CC   arthritis, multiple sclerosis, Sjogren's syndrome, insulin dependent
CC   diabetes mellitus, autoimmune thyroiditis, Crohn's disease, ulcerative
CC   colitis and psoriasis. This sequence encodes the West Nile virus wild-
CC   type capsid protein described as WNVCwt in the disclosure of the
CC   invention
XX
SQ   Sequence 366 BP; 103 A; 81 C; 108 G; 74 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 21;  DB 8;  Length 366;
  Best Local Similarity   100.0%;  Pred. No. 0.57;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCTAAA 21
              |||||||||||||||||||||
Db         30 CCGGGCTGTCAATATGCTAAA 50


RESULT 6
ADQ30647
ID   ADQ30647 standard; DNA; 967 BP.
XX
AC   ADQ30647;
XX
DT   23-SEP-2004  (first entry)
XX
DE   West Nile virus internal diagnosis control sequence.
XX
KW   ss; internal control; West Nile Virus; diagnosis.
XX
OS   West Nile virus.
XX
PN   WO2004055159-A2.
XX
PD   01-JUL-2004.
XX
PF   05-DEC-2003; 2003WO-US038750.
XX
PR   12-DEC-2002; 2002US-0432850P.
PR   20-JUN-2003; 2003US-0480431P.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Shyamala V;
XX
DR   WPI; 2004-488058/46.
XX
PT   New isolated oligonucleotides for accurately diagnosing West Nile virus
PT   infection or for capturing, detecting and quantitating West Nile virus in
PT   blood samples.
XX
PS   Claim 27; SEQ ID NO 17; 56pp; English.
XX
CC   The invention relates to an isolated oligonucleotide not more than 60
CC   nucleotides in length comprising a nucleotide sequence (S1) of at least
CC   10 contiguous nucleotides from any of the 28 nucleotide sequences (e.g.
CC   20, 21 or 23 bp) given in the specification derived from the West Nile
CC   Virus (WNV) genome, a nucleotide sequence (S2) having 90% sequence
CC   identity to the nucleotide sequence of (S1), or complements of (S1) and
CC   (S2). The oligonucleotide further comprises a detectable label at the 5'-
CC   end and/or the 3'-end. The detectable label is a fluorescent label
CC   selected from 6-carboxyfluorescein (6-FAM), tetramethyl rhodamine
CC   (TAMRA), and 2',4',5',7',-tetrachloro-4-7-dichlorofluorescein (TET). The
CC   composition and methods are useful for accurately diagnosing West Nile
CC   virus infection or for capturing, detecting and quantitating West Nile
CC   virus in biological samples, particularly blood samples. This sequence
CC   corresponds to an internal control sequence for the detection of WNV
CC   sequences using the oligonucleotides of the invention.
XX
SQ   Sequence 967 BP; 273 A; 206 C; 272 G; 216 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 21;  DB 12;  Length 967;
  Best Local Similarity   100.0%;  Pred. No. 0.65;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCTAAA 21
              |||||||||||||||||||||
Db        129 CCGGGCTGTCAATATGCTAAA 149


RESULT 7
ADR32078
ID   ADR32078 standard; DNA; 10945 BP.
XX
AC   ADR32078;
XX
```

DT 18-NOV-2004 (first entry)
XX
DE Genomic DNA of a West Nile virus.
XX
KW analysis; target; real time PCR; ds; genomic.
XX
OS West Nile virus.
XX
PN WO2004072230-A2.
XX
PD 26-AUG-2004.
XX
PF 10-FEB-2004; 2004WO-US002012.
XX
PR 10-FEB-2003; 2003US-00361004.
XX
PA (CLEA-) CLEARANT INC.
XX
PI Mckenney K, Gillmeister L, Marlowe K, Armistead D;
XX
DR WPI; 2004-625843/60.
XX
PT Analyzing a target nucleic acid sequence in a biological material by real
PT time PCR using nucleic acid primers that are separated by at least 750
PT nucleic acid residues in the target sequence.
XX
PS Disclosure; SEQ ID NO 5; 96pp; English.
XX
CC The invention relates to a novel method for analysing a target nucleic
CC acid sequence in a biological material. The method comprises adding at
CC least two nucleic acid primers that hybridise under stringent conditions
CC to predetermined nucleic acid sequences of the target nucleic acid
CC sequence that are separated by at least 750 nucleic acid residues,
CC amplifying the target nucleic acid sequence by PCR, and detecting and
CC quantifying the target nucleic acid sequence. The methods and
CC compositions of the present invention are useful for analysing a target
CC nucleic acid sequence in a biological material by real time PCR using
CC nucleic acid primers that are separated by at least 750 nucleic acid
CC residues in the target sequence. This polynucleotide sequence represents
CC the genomic DNA of a West Nile virus used in the target analysis method
CC of the invention.
XX
SQ Sequence 10945 BP; 2999 A; 2457 C; 3143 G; 2346 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 13; Length 10945;
Best Local Similarity 100.0%; Pred. No. 0.93;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCGGGCTGTCAATATGCTAAA 21
||||||||||||||||||||
Db 87 CCGGGCTGTCAATATGCTAAA 107

RESULT 8
ADR67768
ID ADR67768 standard; DNA; 10945 BP.
XX
AC ADR67768;
XX
DT 18-NOV-2004 (first entry)
XX
DE West Nile virus DNA detected by novel detection method.
XX
KW ds; detection; pathogen.
XX
OS West Nile virus.
XX
PN WO2004072231-A2.
XX
PD 26-AUG-2004.
XX
PF 10-FEB-2004; 2004WO-US002013.
XX
PR 10-FEB-2003; 2003US-00361002.
XX
PA (CLEA-) CLEARANT INC.
XX
PI Mckenney K, Gillmeister L, Marlowe K, Armistead D;
XX
DR WPI; 2004-625844/60.
XX
PT Determining level of potentially active biological pathogens in
PT biological material, by adding nucleic acid primer pairs to biological
PT material, amplifying target nucleic acid by PCR, detecting and
PT quantifying target nucleic acid.
XX
PS Disclosure; SEQ ID NO 5; 111pp; English.
XX
CC The invention relates to a method of determining (M1) level of
CC potentially active biological pathogens in biological material, involves
CC adding at least two nucleic acid primer pairs to biological material,
CC amplifying target nucleic acid sequences by PCR, and detecting and
CC quantifying target nucleic acid sequences, where quantity of the nucleic
CC acid sequences is proportional to number of biological pathogens in
CC biological material. (M1) is useful for determining level of potentially
CC active biological pathogens in a biological material such as cells,
CC tissues, blood or blood components, proteins, enzymes, immunoglobulins,
CC botanicals, food, ligaments, tendons, nerves, bone, teeth, skin grafts,
CC bone marrow, heart valves, cartilage, corneas, arteries, veins, organs,
CC lipids, carbohydrates, collagen, chitin and its derivatives, forensic
CC samples, mummified material, human or animal remains, stem cells, islet
CC of Langerhans cells, cells for transplantation, red blood cells, white
CC blood cells or platelets. The biological pathogen is chosen from
CC bacteria, viruses, fungi and single cell parasites. The biological
CC pathogen is chosen from Aspergillus , Candida , Histoplasma ,
CC Saccharomyces , Coccidioides , Cryptococcus , Escherichia , Bacillus ,
CC Campylobacter , Helicobacter , Listeria , Clostridium , Streptococcus ,
CC Enterococcus , Staphylococcus , Brucella , Haemophilus , Salmonella ,
CC Yersinia , Pseudomonas , Serratia , Enterobacter , Klebsiella , Proteus ,
CC Citrobacter , Corynebacterium , Propionibacterium and Coxiella . The
CC biological pathogen is chosen from Adeno-associated virus (AAV),
CC California encephalitis virus, Coronavirus, Coxsackievirus-A,
CC Coxsackievirus-B, Eastern equine encephalitis Virus (EEEV), Echovirus,
CC Hantavirus, Hepatitis A virus (HAV), Hepatitis C virus (HCV), Hepatitis
CC delta virus (HDV), Hepatitis E virus (HEV), Hepatitis G virus (HGV), HIV,
CC Human T-lymphotrophic virus (HTLV), Influenza virus (Flu virus), Measles
CC virus (Rubeola), Mumps virus, Norwalk virus, Parainfluenza virus, Polio
CC virus, Rabies virus, Respiratory Syncytial virus, Rhinovirus, Rubella
CC virus, Saint Louis encephalitis virus, Western equine encephalitis virus
CC (WEEV), Yellow fever virus, Adenovirus, Cytomegalovirus (CMV), Epstein-
CC Barr virus (EBV), Hepatitis B virus (HBV), Herpes simplex virus 1, Herpes
CC simplex virus 2, Molluscum contagiosum, Papilloma virus (HPV), Smallpox
CC virus (Variola), Vaccinia virus, Venezuelan equine encephalitis virus
CC (VEEV), Ebola virus, West Nile virus, Human Parvovirus B19 and Rotavirus.
CC (M1) is useful for determining the effectiveness of a sterilization
CC process applied to a biological material. (M1) is useful in determining
CC whether the biological pathogen is inactive or active. (M1) enables
CC determination of whether the particular biological pathogen is present in
CC a biological material as shown by amplification of first target sequence
CC and whether the biological pathogen is inactive or active. (M1) enables
CC evaluation of the effectiveness of sterilization processes, and
CC determination of both the original level and the residual level of
CC potentially active biological pathogens. This sequence corresponds to a
CC West Nile virus DNA detected by the method of the invention.
XX
SQ Sequence 10945 BP; 2999 A; 2457 C; 3143 G; 2346 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 13; Length 10945;
Best Local Similarity 100.0%; Pred. No. 0.93;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCGGGCTGTCAATATGCTAAA 21
||||||||||||||||||||
Db 87 CCGGGCTGTCAATATGCTAAA 107

```
RESULT 9
ADK13681
ID   ADK13681 standard; DNA; 10962 BP.
XX
AC   ADK13681;
XX
DT   20-MAY-2004  (first entry)
XX
DE   West Nile Virus DNA sequence, SEQ ID 1.
XX
KW   Virucide; Immunostimulant; flavivirus;
KW   envelope protein domain III polypeptide; envelope protein; gene; ss.
XX
OS   West Nile virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             97. .10389
FT                   /*tag=  a
FT                   /product= "West Nile Virus protein"
XX
PN   WO2004016586-A2.
XX
PD   26-FEB-2004.
XX
PF   18-AUG-2003; 2003WO-US025681.
XX
PR   16-AUG-2002; 2002US-0403893P.
PR   06-FEB-2003; 2003US-0445581P.
XX
PA   (TEXA ) UNIV TEXAS SYSTEM.
XX
PI   Barrett A,  Beasley D,  Holbrook M;
XX
DR   WPI; 2004-203756/19.
DR   P-PSDB; ADK13682.
XX
PT   Diagnosing flavivirus infection by contacting a sample from a human or
PT   animal with a flavivirus envelope protein domain III polypeptide, and
PT   detecting formation of a immunocomplex between the envelope protein and
PT   antibodies in the sample.
XX
PS   Disclosure; SEQ ID NO 1; 110pp; English.
XX
CC   The present invention relates to a method for screening for a flavivirus
CC   in a subject or animal host. The method comprises: contacting a sample
CC   from the subject with a composition comprising a flavivirus envelope
CC   protein domain III polypeptide (ADK13683-ADK13701) under conditions that
CC   permit formation of specific immunocomplex between an antibody in the
CC   sample and the envelope protein domain III polypeptide; and detecting
CC   whether a specific immunocomplex is formed. The present sequence is the
CC   coding sequence for West Nile Virus protein, from which E protein
CC   envelope protein domain III polypeptide (ADK13683) is derived.
XX
SQ   Sequence 10962 BP; 2997 A; 2497 C; 3100 G; 2368 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 21;  DB 12;  Length 10962;
  Best Local Similarity   100.0%;  Pred. No. 0.93;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCTAAA 21
              |||||||||||||||||||||
Db        129 CCGGGCTGTCAATATGCTAAA 149


RESULT 10
ADN98022
ID   ADN98022 standard; DNA; 10975 BP.
XX
AC   ADN98022;
XX
DT   29-JUL-2004  (first entry)
XX
DE   West Nile Virus isolate 2741 complete genome sequence.
XX
KW   ds; West Nile Virus; envelope protein; glycoprotein E; flavivirus;
KW   Japanese encephalitis virus; Dengue virus; St Louis encephalitis virus.
XX
OS   West Nile virus.
XX
PN   WO2004040263-A2.
XX
PD   13-MAY-2004.
XX
PF   31-OCT-2003; 2003WO-US034823.
XX
PR   31-OCT-2002; 2002US-0422755P.
PR   06-JUN-2003; 2003US-0476513P.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Wong SJ,  Pei-Yong S;
XX
DR   WPI; 2004-400223/37.
DR   GENBANK; AF206518.
XX
PT   New diagnostic kit comprising West Nile Virus (WNV) envelope protein
PT   reactive with antibody against WNV and cross-reactive with antibody
PT   against a flavivirus, useful in diagnosing flavivirus infection caused by
PT   DENV, WNV, JEV or SLEV.
XX
PS   Disclosure; Fig 37; 212pp; English.
XX
CC   The invention relates to a diagnostic kit comprising at least one
CC   isolated and purified polypeptide comprising a West Nile Virus (WNV)
CC   envelope (E) protein or its immunogenic fragment having a native
CC   conformation or non-denatured structure and that is reactive with
CC   antibodies against WNV and cross-reactive with antibodies against a
CC   flavivirus. The diagnostic kit is useful in diagnosing flavivirus
CC   infection caused by DENV, WNV, JEV or SLEV. This sequence corresponds to
CC   the complete nucleotide sequence of the WNV isolate 2741.
XX
SQ   Sequence 10975 BP; 3007 A; 2460 C; 3149 G; 2359 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 21;  DB 12;  Length 10975;
  Best Local Similarity   100.0%;  Pred. No. 0.93;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCTAAA 21
              |||||||||||||||||||||
Db        111 CCGGGCTGTCAATATGCTAAA 131


RESULT 11
ABZ68481
ID   ABZ68481 standard; DNA; 11029 BP.
XX
AC   ABZ68481;
XX
DT   22-APR-2003  (first entry)
XX
DE   Nucleotide sequence of the genome of West Nile virus IS-98-ST1.
XX
KW   WNV; IS-98-ST1; Flavivirus; infection; encephalitis; gene; ss.
XX
OS   West nile virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             97. .10397
FT                   /*tag=  a
FT                   /product= "polyprotein"
XX
PN   WO200281511-A1.
XX
PD   17-OCT-2002.
XX
PF   04-APR-2002; 2002WO-FR001168.
```

```
XX
PR   04-APR-2001; 2001FR-00004599.
PR   06-SEP-2001; 2001FR-00011525.
XX
PA   (INSP ) INST PASTEUR.
PA   (KIMR-) KIMRON VETERINARY INST.
XX
PI   Despres P,  Deubel V,  Guenet J,  Drouet M,  Malkinson M,  Banet C;
PI   Frenkiel M,  Courageot M,  Coulibaly F,  Catteau A,  Flamand M,  Weber P;
PI   Ceccaldi P;
XX
DR   WPI; 2003-058498/05.
DR   P-PSDB; ABP70647.
XX
PT   New neurovirulent strain of West Nile virus, useful in diagnosis and
PT   screening for antiviral agents, also related nucleic acids, proteins and
PT   antibodies.
XX
PS   Claim 1; Page 34-49; 68pp; French.
XX
CC   The present sequence represents the genome of a strain of West Nile virus
CC   (WNV), designated IS-98-ST1. This strain is a neuroinvasive and
CC   neurovirulent strain of WNV. Polynucleotides and polypeptides derived
CC   from the IS-98-ST1 genome are useful for diagnosis and prognosis of
CC   Flavivirus infection, specifically WNV-mediated encephalitis. They are
CC   also useful to raise specific antibodies, for recombinant expression of
CC   WNV proteins or peptides (for diagnosis, production of antibodies and
CC   identification of specific binding partners in cells), for identifying
CC   cellular genes implicated in resistance to viral infection, and for
CC   screening for anti-Flavivirus agents
XX
SQ   Sequence 11029 BP; 3019 A; 2471 C; 3167 G; 2372 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 21;  DB 8;  Length 11029;
  Best Local Similarity   100.0%;  Pred. No. 0.93;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCTAAA 21
              |||||||||||||||||||||
Db        129 CCGGGCTGTCAATATGCTAAA 149


RESULT 12
ABV74821
ID   ABV74821 standard; DNA; 11029 BP.
XX
AC   ABV74821;
XX
DT   28-MAR-2003  (first entry)
XX
DE   West Nile virus strain NY99-flamingo 382-99 complete genome.
XX
KW   Virucide; hepatotropic; antiinflammatory; antiviral; OAS;
KW   2'-5'-oligoadenylate synthase; Flavivirus infection; gene; ss.
XX
OS   West Nile Virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             97. .10398
FT                   /*tag=  a
FT                   /product= "West Nile Virus protein"
XX
PN   WO200281741-A2.
XX
PD   17-OCT-2002.
XX
PF   04-APR-2002; 2002WO-FR001169.
XX
PR   04-APR-2001; 2001FR-00004598.
XX
PA   (INSP ) INST PASTEUR.
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Guenet J,  Mashimo T,  Simon-Chazottes D,  Montagutelli X;
PI   Frenkiel M,  Despres P,  Deubel V,  Bonhomme F,  Lucas M;
XX
DR   WPI; 2003-058566/05.
DR   P-PSDB; ABB98821.
XX
PT   Identifying stimulators of oligoadenylate synthase family genes, useful
PT   as antiviral agents against Flavivirus, also mutated genes responsible
PT   for sensitivity to virus.
XX
PS   Example 1; Page 52-67; 93pp; French.
XX
CC   The present invention relates to a method for identifying compounds (I)
CC   that can stimulate a gene of the OAS (2'-5'-oligoadenylate synthase)
CC   family. The method comprises: (a) inducing expression of the OAS gene in
CC   a culture of cells from a non-human mammal (Flvr/Flvr or Flvr/Flvs;
CC   indicating resistance or sensitivity to Flavivirus infection); (b)
CC   treating cells with test compound; and (c) measuring activity of OAS gene
CC   relative to a control. (I) are potentially useful as antiviral agents for
CC   treating infections by Flaviviruses (e.g. hepatitis C; dengue; yellow
CC   fever and various forms of encephalitis). Genomic OAS DNA and derived
CC   cDNA, also the encoded proteins, are useful: (a) for treating Flavivirus
CC   infection; (b) in screening for anti-flavivirus agents, and (c) for
CC   evaluating sensitivity of subjects to Flavivirus infection and their
CC   likely response to interferon treatment, e.g. to identify patients at
CC   risk of developing severe forms of such infections. The present sequence
CC   is West Nile Virus strain NY99-flamingo 382-99 (IS-98-STI) complete
CC   genome, which was used in an example from the invention. West Nile Virus
CC   is one such Flavivirus
XX
SQ   Sequence 11029 BP; 3019 A; 2471 C; 3167 G; 2372 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 21;  DB 10;  Length 11029;
  Best Local Similarity   100.0%;  Pred. No. 0.93;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCTAAA 21
              |||||||||||||||||||||
Db        129 CCGGGCTGTCAATATGCTAAA 149


RESULT 13
ADN98023
ID   ADN98023 standard; DNA; 11029 BP.
XX
AC   ADN98023;
XX
DT   29-JUL-2004  (first entry)
XX
DE   West Nile Virus isolate 3356 complete genome sequence.
XX
KW   ds; West Nile Virus; envelope protein; glycoprotein E; flavivirus;
KW   Japanese encephalitis virus; Dengue virus; St Louis encephalitis virus.
XX
OS   West Nile virus.
XX
PN   WO2004040263-A2.
XX
PD   13-MAY-2004.
XX
PF   31-OCT-2003; 2003WO-US034823.
XX
PR   31-OCT-2002; 2002US-0422755P.
PR   06-JUN-2003; 2003US-0476513P.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Wong SJ,  Pei-Yong S;
XX
DR   WPI; 2004-400223/37.
DR   GENBANK; AF404756.
XX
PT   New diagnostic kit comprising West Nile Virus (WNV) envelope protein
```

```
PT   reactive with antibody against WNV and cross-reactive with antibody
PT   against a flavivirus, useful in diagnosing flavivirus infection caused by
PT   DENV, WNV, JEV or SLEV.
XX
PS   Disclosure; Fig 38; 212pp; English.
XX
CC   The invention relates to a diagnostic kit comprising at least one
CC   isolated and purified polypeptide comprising a West Nile Virus (WNV)
CC   envelope (E) protein or its immunogenic fragment having a native
CC   conformation or non-denatured structure and that is reactive with
CC   antibodies against WNV and cross-reactive with antibodies against a
CC   flavivirus. The diagnostic kit is useful in diagnosing flavivirus
CC   infection caused by DENV, WNV, JEV or SLEV. This sequence corresponds to
CC   the complete nucleotide sequence of the WNV isolate 3356.
XX
SQ   Sequence 11029 BP; 3017 A; 2466 C; 3172 G; 2374 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 21;  DB 12;  Length 11029;
  Best Local Similarity   100.0%;  Pred. No. 0.93;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCTAAA 21
              |||||||||||||||||||||
Db        129 CCGGGCTGTCAATATGCTAAA 149


RESULT 14
ADN36695/c
ID   ADN36695 standard; DNA; 21 BP.
XX
AC   ADN36695;
XX
DT   15-JUL-2004  (first entry)
XX
DE   West Nile virus detection-related oligonucleotide probe SeqID17.
XX
KW   hybridisation assay probe; nucleic acid detection;
KW   target-complementary sequence; flavivirus; West Nile virus; WNV;
KW   RNA virus; infection; meningitis; encephalitis;
KW   high throughput screening; probe; ss.
XX
OS   West Nile virus.
XX
PN   WO2004036190-A2.
XX
PD   29-APR-2004.
XX
PF   10-OCT-2003; 2003WO-US033639.
XX
PR   16-OCT-2002; 2002US-0418891P.
PR   25-NOV-2002; 2002US-0429006P.
PR   24-FEB-2003; 2003US-0449810P.
XX
PA   (GENP-) GEN-PROBE INC.
XX
PI   Linnen JM,  Pollner RB,  Wu W,  Dennis GG,  Darby PM;
XX
DR   WPI; 2004-389590/36.
XX
PT   New hybridization assay probe comprising target-complementary sequence of
PT   bases, useful in detecting flavivirus, e.g. West Nile virus.
XX
PS   Disclosure; SEQ ID NO 17; 135pp; English.
XX
CC   This invention relates to a novel hybridisation assay probe, for
CC   detecting a nucleic acid, which is a probe sequence that comprises a
CC   target-complementary sequence of bases, and optionally one or more base
CC   sequences that are not complementary to the nucleic acid that is to be
CC   detected. The hybridisation assay probes and the kits are useful in
CC   detecting and amplifying a target nucleic acid sequence, for example
CC   flavivirus like West Nile virus, that may be present in a biological
CC   sample. West Nile virus (WNV) is an RNA virus that primarily infects
CC   birds and culex mosquitos, with humans and horses serving as incidental
CC   hosts. Infection of humans can lead to meningitis or encephalitis. The
CC   invention may allow for accurate and efficient high throughput screening.
CC   The present sequence is that of an oligonucleotide probe which is related
CC   to the invention.
XX
SQ   Sequence 21 BP; 5 A; 5 C; 4 G; 7 T; 0 U; 0 Other;

  Query Match             90.5%;  Score 19;  DB 12;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 4.4;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 GGGCTGTCAATATGCTAAA 21
              |||||||||||||||||||
Db         21 GGGCTGTCAATATGCTAAA 3


RESULT 15
ADN36823/c
ID   ADN36823 standard; RNA; 24 BP.
XX
AC   ADN36823;
XX
DT   15-JUL-2004  (first entry)
XX
DE   West Nile virus detection-related oligonucleotide probe SeqID145.
XX
KW   hybridisation assay probe; nucleic acid detection;
KW   target-complementary sequence; flavivirus; West Nile virus; WNV;
KW   RNA virus; infection; meningitis; encephalitis;
KW   high throughput screening; probe; ss.
XX
OS   West Nile virus.
XX
FH   Key             Location/Qualifiers
FT   modified_base   1. .24
FT                   /*tag=  a
FT                   /mod_base= OTHER
FT                   /note= "OTHER= 2'-methoxyethoxy (2'-MOE) nucleotides"
XX
PN   WO2004036190-A2.
XX
PD   29-APR-2004.
XX
PF   10-OCT-2003; 2003WO-US033639.
XX
PR   16-OCT-2002; 2002US-0418891P.
PR   25-NOV-2002; 2002US-0429006P.
PR   24-FEB-2003; 2003US-0449810P.
XX
PA   (GENP-) GEN-PROBE INC.
XX
PI   Linnen JM,  Pollner RB,  Wu W,  Dennis GG,  Darby PM;
XX
DR   WPI; 2004-389590/36.
XX
PT   New hybridization assay probe comprising target-complementary sequence of
PT   bases, useful in detecting flavivirus, e.g. West Nile virus.
XX
PS   Example 1; SEQ ID NO 145; 135pp; English.
XX
CC   This invention relates to a novel hybridisation assay probe, for
CC   detecting a nucleic acid, which is a probe sequence that comprises a
CC   target-complementary sequence of bases, and optionally one or more base
CC   sequences that are not complementary to the nucleic acid that is to be
CC   detected. The hybridisation assay probes and the kits are useful in
CC   detecting and amplifying a target nucleic acid sequence, for example
CC   flavivirus like West Nile virus, that may be present in a biological
CC   sample. West Nile virus (WNV) is an RNA virus that primarily infects
CC   birds and culex mosquitos, with humans and horses serving as incidental
CC   hosts. Infection of humans can lead to meningitis or encephalitis. The
CC   invention may allow for accurate and efficient high throughput screening.
CC   The present sequence is that of an oligonucleotide probe which is related
CC   to the invention.
```

```
XX
SQ   Sequence 24 BP; 5 A; 7 C; 5 G; 0 T; 7 U; 0 Other;

  Query Match             90.5%;  Score 19;  DB 12;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 4.4;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 GGGCTGTCAATATGCTAAA 21
              |||||||||||||||||||
Db         24 GGGCTGTCAATATGCTAAA 6


RESULT 16
ADN36707/c
ID   ADN36707 standard; DNA; 48 BP.
XX
AC   ADN36707;
XX
DT   15-JUL-2004  (first entry)
XX
DE   West Nile virus detection-related oligonucleotide probe SeqID29.
XX
KW   hybridisation assay probe; nucleic acid detection;
KW   target-complementary sequence; flavivirus; West Nile virus; WNV;
KW   RNA virus; infection; meningitis; encephalitis;
KW   high throughput screening; probe; ss.
XX
OS   West Nile virus.
OS   Enterobacteria phage T7.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1. .27
FT                   /*tag=  a
FT                   /note= "T7 promoter sequence"
FT   misc_feature    28. .48
FT                   /*tag=  b
FT                   /note= "WNV-complimentary sequence"
XX
PN   WO2004036190-A2.
XX
PD   29-APR-2004.
XX
PF   10-OCT-2003; 2003WO-US033639.
XX
PR   16-OCT-2002; 2002US-0418891P.
PR   25-NOV-2002; 2002US-0429006P.
PR   24-FEB-2003; 2003US-0449810P.
XX
PA   (GENP-) GEN-PROBE INC.
XX
PI   Linnen JM,  Pollner RB,  Wu W,  Dennis GG,  Darby PM;
XX
DR   WPI; 2004-389590/36.
XX
PT   New hybridization assay probe comprising target-complementary sequence of
PT   bases, useful in detecting flavivirus, e.g. West Nile virus.
XX
PS   Disclosure; SEQ ID NO 29; 135pp; English.
XX
CC   This invention relates to a novel hybridisation assay probe, for
CC   detecting a nucleic acid, which is a probe sequence that comprises a
CC   target-complementary sequence of bases, and optionally one or more base
CC   sequences that are not complementary to the nucleic acid that is to be
CC   detected. The hybridisation assay probes and the kits are useful in
CC   detecting and amplifying a target nucleic acid sequence, for example
CC   flavivirus like West Nile virus, that may be present in a biological
CC   sample. West Nile virus (WNV) is an RNA virus that primarily infects
CC   birds and culex mosquitos, with humans and horses serving as incidental
CC   hosts. Infection of humans can lead to meningitis or encephalitis. The
CC   invention may allow for accurate and efficient high throughput screening.
CC   The present sequence is that of an oligonucleotide probe which is related
CC   to the invention.
XX
SQ   Sequence 48 BP; 16 A; 9 C; 9 G; 14 T; 0 U; 0 Other;

  Query Match             90.5%;  Score 19;  DB 12;  Length 48;
  Best Local Similarity   100.0%;  Pred. No. 4.9;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 GGGCTGTCAATATGCTAAA 21
              |||||||||||||||||||
Db         48 GGGCTGTCAATATGCTAAA 30


RESULT 17
ADN36694/c
ID   ADN36694 standard; DNA; 69 BP.
XX
AC   ADN36694;
XX
DT   15-JUL-2004  (first entry)
XX
DE   West Nile virus detection-related oligonucleotide probe SeqID16.
XX
KW   hybridisation assay probe; nucleic acid detection;
KW   target-complementary sequence; flavivirus; West Nile virus; WNV;
KW   RNA virus; infection; meningitis; encephalitis;
KW   high throughput screening; probe; ss.
XX
OS   West Nile virus.
XX
PN   WO2004036190-A2.
XX
PD   29-APR-2004.
XX
PF   10-OCT-2003; 2003WO-US033639.
XX
PR   16-OCT-2002; 2002US-0418891P.
PR   25-NOV-2002; 2002US-0429006P.
PR   24-FEB-2003; 2003US-0449810P.
XX
PA   (GENP-) GEN-PROBE INC.
XX
PI   Linnen JM,  Pollner RB,  Wu W,  Dennis GG,  Darby PM;
XX
DR   WPI; 2004-389590/36.
XX
PT   New hybridization assay probe comprising target-complementary sequence of
PT   bases, useful in detecting flavivirus, e.g. West Nile virus.
XX
PS   Claim 68; SEQ ID NO 16; 135pp; English.
XX
CC   This invention relates to a novel hybridisation assay probe, for
CC   detecting a nucleic acid, which is a probe sequence that comprises a
CC   target-complementary sequence of bases, and optionally one or more base
CC   sequences that are not complementary to the nucleic acid that is to be
CC   detected. The hybridisation assay probes and the kits are useful in
CC   detecting and amplifying a target nucleic acid sequence, for example
CC   flavivirus like West Nile virus, that may be present in a biological
CC   sample. West Nile virus (WNV) is an RNA virus that primarily infects
CC   birds and culex mosquitos, with humans and horses serving as incidental
CC   hosts. Infection of humans can lead to meningitis or encephalitis. The
CC   invention may allow for accurate and efficient high throughput screening.
CC   The present sequence is that of an oligonucleotide probe which is related
CC   to the invention.
XX
SQ   Sequence 69 BP; 18 A; 22 C; 14 G; 15 T; 0 U; 0 Other;

  Query Match             90.5%;  Score 19;  DB 12;  Length 69;
  Best Local Similarity   100.0%;  Pred. No. 5.2;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 GGGCTGTCAATATGCTAAA 21
              |||||||||||||||||||
Db         69 GGGCTGTCAATATGCTAAA 51
```

```
RESULT 18
ADN36696/c
ID   ADN36696 standard; DNA; 20 BP.
XX
AC   ADN36696;
XX
DT   15-JUL-2004  (first entry)
XX
DE   West Nile virus detection-related oligonucleotide probe SeqID18.
XX
KW   hybridisation assay probe; nucleic acid detection;
KW   target-complementary sequence; flavivirus; West Nile virus; WNV;
KW   RNA virus; infection; meningitis; encephalitis;
KW   high throughput screening; probe; ss.
XX
OS   West Nile virus.
XX
PN   WO2004036190-A2.
XX
PD   29-APR-2004.
XX
PF   10-OCT-2003; 2003WO-US033639.
XX
PR   16-OCT-2002; 2002US-0418891P.
PR   25-NOV-2002; 2002US-0429006P.
PR   24-FEB-2003; 2003US-0449810P.
XX
PA   (GENP-) GEN-PROBE INC.
XX
PI   Linnen JM,  Pollner RB,  Wu W,  Dennis GG,  Darby PM;
XX
DR   WPI; 2004-389590/36.
XX
PT   New hybridization assay probe comprising target-complementary sequence of
PT   bases, useful in detecting flavivirus, e.g. West Nile virus.
XX
PS   Disclosure; SEQ ID NO 18; 135pp; English.
XX
CC   This invention relates to a novel hybridisation assay probe, for
CC   detecting a nucleic acid, which is a probe sequence that comprises a
CC   target-complementary sequence of bases, and optionally one or more base
CC   sequences that are not complementary to the nucleic acid that is to be
CC   detected. The hybridisation assay probes and the kits are useful in
CC   detecting and amplifying a target nucleic acid sequence, for example
CC   flavivirus like West Nile virus, that may be present in a biological
CC   sample. West Nile virus (WNV) is an RNA virus that primarily infects
CC   birds and culex mosquitos, with humans and horses serving as incidental
CC   hosts. Infection of humans can lead to meningitis or encephalitis. The
CC   invention may allow for accurate and efficient high throughput screening.
CC   The present sequence is that of an oligonucleotide probe which is related
CC   to the invention.
XX
SQ   Sequence 20 BP; 5 A; 4 C; 4 G; 7 T; 0 U; 0 Other;

  Query Match             85.7%;  Score 18;  DB 12;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 GGCTGTCAATATGCTAAA 21
              ||||||||||||||||||
Db         20 GGCTGTCAATATGCTAAA 3


RESULT 19
ADN36708/c
ID   ADN36708 standard; DNA; 47 BP.
XX
AC   ADN36708;
XX
DT   15-JUL-2004  (first entry)
XX
DE   West Nile virus detection-related oligonucleotide probe SeqID30.
XX
KW   hybridisation assay probe; nucleic acid detection;
KW   target-complementary sequence; flavivirus; West Nile virus; WNV;
KW   RNA virus; infection; meningitis; encephalitis;
KW   high throughput screening; probe; ss.
XX
OS   West Nile virus.
OS   Enterobacteria phage T7.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1. .27
FT                   /*tag=  a
FT                   /note= "T7 promoter sequence"
FT   misc_feature    28. .47
FT                   /*tag=  b
FT                   /note= "WNV-complimentary sequence"
XX
PN   WO2004036190-A2.
XX
PD   29-APR-2004.
XX
PF   10-OCT-2003; 2003WO-US033639.
XX
PR   16-OCT-2002; 2002US-0418891P.
PR   25-NOV-2002; 2002US-0429006P.
PR   24-FEB-2003; 2003US-0449810P.
XX
PA   (GENP-) GEN-PROBE INC.
XX
PI   Linnen JM,  Pollner RB,  Wu W,  Dennis GG,  Darby PM;
XX
DR   WPI; 2004-389590/36.
XX
PT   New hybridization assay probe comprising target-complementary sequence of
PT   bases, useful in detecting flavivirus, e.g. West Nile virus.
XX
PS   Disclosure; SEQ ID NO 30; 135pp; English.
XX
CC   This invention relates to a novel hybridisation assay probe, for
CC   detecting a nucleic acid, which is a probe sequence that comprises a
CC   target-complementary sequence of bases, and optionally one or more base
CC   sequences that are not complementary to the nucleic acid that is to be
CC   detected. The hybridisation assay probes and the kits are useful in
CC   detecting and amplifying a target nucleic acid sequence, for example
CC   flavivirus like West Nile virus, that may be present in a biological
CC   sample. West Nile virus (WNV) is an RNA virus that primarily infects
CC   birds and culex mosquitos, with humans and horses serving as incidental
CC   hosts. Infection of humans can lead to meningitis or encephalitis. The
CC   invention may allow for accurate and efficient high throughput screening.
CC   The present sequence is that of an oligonucleotide probe which is related
CC   to the invention.
XX
SQ   Sequence 47 BP; 16 A; 8 C; 9 G; 14 T; 0 U; 0 Other;

  Query Match             85.7%;  Score 18;  DB 12;  Length 47;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 GGCTGTCAATATGCTAAA 21
              ||||||||||||||||||
Db         47 GGCTGTCAATATGCTAAA 30


RESULT 20
AAQ22767
ID   AAQ22767 standard; DNA; 4512 BP.
XX
AC   AAQ22767;
XX
DT   12-AUG-1992  (first entry)
XX
DE   JEV Nakayama strain prM, E, NS1, NS2A, NS2B and C coding regions.
XX
```

KW Japanese Encephalitis Virus; vaccinia virus donor; plasmid pDr20; ss.
XX
OS Japanese encephalitis virus.
XX
PN WO9203545-A.
XX
PD 05-MAR-1992.
XX
PF 05-AUG-1991; 91WO-US005816.
XX
PR 15-AUG-1990; 90US-00567960.
PR 06-JUN-1991; 91US-00711429.
PR 13-JUN-1991; 91US-00714687.
PR 17-JUL-1991; 91US-00729800.
PR 05-AUG-1991; 91WO-UO005816.
XX
PA (VIRO-) VIROGENETICS CORP.
XX
PI Paoletti E, Pinc, Pinc, Pincus SE;
XX
DR WPI; 1992-096889/12.
XX
PT Recombinant pox-virus e.g. vaccinia, fowl-pox and canary-pox virus -
PT contg. DNA from flavi-virus e.g. Japanese encephalitis and yellow fever
PT virus, used as vaccine.
XX
PS Example 9; Fig 17; 117pp; English.
XX
CC cDNA was prepared from genomic virion RNA obtained from suspension
CC cultures of C6/36 cells infected with a passage 55 suckling mouse brain
CC stock of the Nagayama strain of JEV. EcoRI linkers were ligated to the
CC cDNA fragments for cloning into pBR322. Recombinant plasmids were
CC transformed into E.coli DH5 cells. Plasmid pC20 contained 81 non-coding
CC nucleotides and the coding regions for C and prM. Sequence AAQ22767 is
CC that of the C coding region of pC20, combined with an updated sequence of
CC the prM, E, NS1, NS2A and NS2B coding regions of the Nagayama strain. The
CC sequence begins at the C protein Met initiation codon. A subfragment of
CC pC20 was cloned into pUC18 to give pDr20. This plasmid was then used in
CC the construction of novel recombinants JEV24, JEV27, JEV33 and JEV34.
CC These were transfected into vP410 infected cells to generate vP825,
CC vP829, vP857 and vP864, respectively
XX
SQ Sequence 4512 BP; 1192 A; 1055 C; 1253 G; 1012 T; 0 U; 0 Other;

Query Match 84.8%; Score 17.8; DB 2; Length 4512;
Best Local Similarity 90.5%; Pred. No. 41;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCGGGCTGTCAATATGCTAAA 21
||||||| |||||||||| ||
Db 33 CCGGGCTATCAATATGCTGAA 53

RESULT 21
ADO07431
ID ADO07431 standard; DNA; 10818 BP.
XX
AC ADO07431;
XX
DT 15-JUL-2004 (first entry)
XX
DE Japanese Encephalitis virus JEV coding sequence SEQ ID NO: 9.
XX
KW antiinflammatory; neuroprotective; gene therapy;
KW Japanese Encephalitis virus; JEV; ds; gene; vaccine;
KW japanese encephalitis.
XX
OS Japanese encephalitis virus.
XX
PN WO2004033690-A1.
XX
PD 22-APR-2004.
XX
PF 09-OCT-2003; 2003WO-KR002081.
XX
PR 09-OCT-2002; 2002KR-00061589.
XX
PA (CIDC-) CID CO LTD.
PA (LEES/) LEE S H.
XX
PI Lee SH, Lee Y, Yun S;
XX
DR WPI; 2004-340933/31.
XX
PT New Japanese encephalitis virus genomic RNA, useful in developing
PT vaccines for and in diagnosing and treating Japanese encephalitis.
XX
PS Example 2; Page 145-152; 265pp; English.
XX
CC The present invention relates to a genomic RNA of the Korean Japanese
CC Encephalitis Virus (JEV) isolate, composed of a 5' and 3' nontranslated
CC region (NTR) and a single polypeptide coding region. The JEV genomic RNA,
CC JEV cDNA and reagents are useful in developing vaccines for and in
CC diagnosing and treating Japanese encephalitis. The present sequence is a
CC sequence of the invention.
XX
SQ Sequence 10818 BP; 2991 A; 2491 C; 3075 G; 2261 T; 0 U; 0 Other;

Query Match 84.8%; Score 17.8; DB 12; Length 10818;
Best Local Similarity 90.5%; Pred. No. 47;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCGGGCTGTCAATATGCTAAA 21
||||||| |||||||||| ||
Db 78 CCGGGCTATCAATATGCTGAA 98

RESULT 22
ADO07437
ID ADO07437 standard; DNA; 10968 BP.
XX
AC ADO07437;
XX
DT 15-JUL-2004 (first entry)
XX
DE Japanese Encephalitis virus JEV coding sequence SEQ ID NO: 15.
XX
KW antiinflammatory; neuroprotective; gene therapy;
KW Japanese Encephalitis virus; JEV; ds; gene; vaccine;
KW japanese encephalitis.
XX
OS Japanese encephalitis virus.
XX
PN WO2004033690-A1.
XX
PD 22-APR-2004.
XX
PF 09-OCT-2003; 2003WO-KR002081.
XX
PR 09-OCT-2002; 2002KR-00061589.
XX
PA (CIDC-) CID CO LTD.
PA (LEES/) LEE S H.
XX
PI Lee SH, Lee Y, Yun S;
XX
DR WPI; 2004-340933/31.
XX
PT New Japanese encephalitis virus genomic RNA, useful in developing
PT vaccines for and in diagnosing and treating Japanese encephalitis.
XX
PS Claim 3; Page 154-161; 265pp; English.
XX
CC The present invention relates to a genomic RNA of the Korean Japanese
CC Encephalitis Virus (JEV) isolate, composed of a 5' and 3' nontranslated
CC region (NTR) and a single polypeptide coding region. The JEV genomic RNA,

```
CC   JEV cDNA and reagents are useful in developing vaccines for and in
CC   diagnosing and treating Japanese encephalitis. The present sequence is a
CC   sequence of the invention.
XX
SQ   Sequence 10968 BP; 3032 A; 2518 C; 3117 G; 2301 T; 0 U; 0 Other;

  Query Match             84.8%;  Score 17.8;  DB 12;  Length 10968;
  Best Local Similarity   90.5%;  Pred. No. 47;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCTAAA 21
              ||||||| |||||||||| ||
Db        128 CCGGGCTATCAATATGCTGAA 148


RESULT 23
ABL50890
ID   ABL50890 standard; RNA; 10976 BP.
XX
AC   ABL50890;
XX
DT   23-JUN-2002  (first entry)
XX
DE   Japanese encephalitis virus strain SA14-14-2 polyprotein RNA sequence.
XX
KW   Japanese encephalitis virus strain SA14-14-2; polyprotein; vaccine; gene;
KW   ss.
XX
OS   Japanese encephalitis virus.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1. .95
FT                   /*tag=  a
FT   CDS             96. .10394
FT                   /*tag=  b
FT                   /product= "polyprotein"
FT   mat_peptide     96. .476
FT                   /*tag=  c
FT                   /product= "capsid protein"
FT                   /note= "C"
FT   mat_peptide     477. .977
FT                   /*tag=  d
FT                   /product= "premembrane/membrane protein"
FT                   /note= "PrM/M"
FT   mat_peptide     978. .2477
FT                   /*tag=  e
FT                   /product= "envelope glycoprotein"
FT                   /note= "E"
FT   mat_peptide     2478. .3722
FT                   /*tag=  f
FT                   /product= "nonstructural protein 1"
FT                   /note= "NS1"
FT   mat_peptide     3723. .4214
FT                   /*tag=  g
FT                   /product= "nonstructural protein 2A"
FT                   /note= "NS2A"
FT   mat_peptide     4215. .4607
FT                   /*tag=  h
FT                   /product= "nonstructural protein 2B"
FT                   /note= "NS2B"
FT   mat_peptide     4608. .6464
FT                   /*tag=  i
FT                   /product= "nonstructural protein 3"
FT                   /note= "NS3"
FT   mat_peptide     6465. .7265
FT                   /*tag=  j
FT                   /product= "nonstructural protein 4A"
FT                   /note= "NS4A"
FT   mat_peptide     7266. .10391
FT                   /*tag=  k
FT                   /product= "nonstructural protein 4B"
FT                   /note= "NS4B"
FT   mat_peptide     10395. .10976
FT                   /*tag=  l
FT                   /product= "nonstructural protein 5"
FT                   /note= "NS5"
FT   3'UTR           10395. .10976
FT                   /*tag=  m
XX
PN   KR99038070-A.
XX
PD   05-JUN-1999.
XX
PF   03-NOV-1997;   97KR-00057701.
XX
PR   03-NOV-1997;   97KR-00057701.
XX
PA   (KIMC/) KIM C R.
XX
PI   Yang HJ;
XX
DR   WPI; 2000-383865/33.
DR   P-PSDB; ABB07037.
XX
PT   Attenuated Japanese encephalitis virus and process for preparing vaccine
PT   containing the same.
XX
PS   Claim 3; Page 17-19; 19pp; Korean.
XX
CC   The present invention describes an attenuated Japanese encephalitis virus
CC   and a process for preparing a vaccine containing the same attenuated
CC   Japanese encephalitis virus. The present sequence encodes a polyprotein
CC   from Japanese encephalitis virus strain SA14-14-2, which is given in the
CC   present invention
XX
SQ   Sequence 10976 BP; 3039 A; 2516 C; 3112 G; 0 T; 2309 U; 0 Other;

  Query Match             84.8%;  Score 17.8;  DB 3;  Length 10976;
  Best Local Similarity   66.7%;  Pred. No. 47;
  Matches   14;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCTAAA 21
              ||||||: :|||:|:||: ||
Db        128 CCGGGCUAUCAAUAUGCUGAA 148


RESULT 24
ADO07466
ID   ADO07466 standard; DNA; 18563 BP.
XX
AC   ADO07466;
XX
DT   15-JUL-2004  (first entry)
XX
DE   Japanese Encephalitis virus JEV coding sequence SEQ ID NO: 44.
XX
KW   antiinflammatory; neuroprotective; gene therapy;
KW   Japanese Encephalitis virus; JEV; ds; gene; vaccine;
KW   japanese encephalitis.
XX
OS   Japanese encephalitis virus.
XX
PN   WO2004033690-A1.
XX
PD   22-APR-2004.
XX
PF   09-OCT-2003; 2003WO-KR002081.
XX
PR   09-OCT-2002; 2002KR-00061589.
XX
PA   (CIDC-) CID CO LTD.
PA   (LEES/) LEE S H.
XX
PI   Lee SH,  Lee Y,  Yun S;
XX
DR   WPI; 2004-340933/31.
```

XX
PT New Japanese encephalitis virus genomic RNA, useful in developing
PT vaccines for and in diagnosing and treating Japanese encephalitis.
XX
PS Claim 12; Page 193-206; 265pp; English.
XX
CC The present invention relates to a genomic RNA of the Korean Japanese
CC Encephalitis Virus (JEV) isolate, composed of a 5' and 3' nontranslated
CC region (NTR) and a single polypeptide coding region. The JEV genomic RNA,
CC JEV cDNA and reagents are useful in developing vaccines for and in
CC diagnosing and treating Japanese encephalitis. The present sequence is a
CC sequence of the invention.
XX
SQ Sequence 18563 BP; 4943 A; 4211 C; 4929 G; 4480 T; 0 U; 0 Other;

Query Match 84.8%; Score 17.8; DB 12; Length 18563;
Best Local Similarity 90.5%; Pred. No. 51;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCGGGCTGTCAATATGCTAAA 21
||||||| |||||||||| ||
Db 128 CCGGGCTATCAATATGCTGAA 148

RESULT 25
ADO07465
ID ADO07465 standard; DNA; 18563 BP.
XX
AC ADO07465;
XX
DT 15-JUL-2004 (first entry)
XX
DE Japanese Encephalitis virus JEV coding sequence SEQ ID NO: 43.
XX
KW antiinflammatory; neuroprotective; gene therapy;
KW Japanese Encephalitis virus; JEV; ds; gene; vaccine;
KW japanese encephalitis.
XX
OS Japanese encephalitis virus.
XX
PN WO2004033690-A1.
XX
PD 22-APR-2004.
XX
PF 09-OCT-2003; 2003WO-KR002081.
XX
PR 09-OCT-2002; 2002KR-00061589.
XX
PA (CIDC-) CID CO LTD.
PA (LEES/) LEE S H.
XX
PI Lee SH, Lee Y, Yun S;
XX
DR WPI; 2004-340933/31.
XX
PT New Japanese encephalitis virus genomic RNA, useful in developing
PT vaccines for and in diagnosing and treating Japanese encephalitis.
XX
PS Claim 12; Page 180-193; 265pp; English.
XX
CC The present invention relates to a genomic RNA of the Korean Japanese
CC Encephalitis Virus (JEV) isolate, composed of a 5' and 3' nontranslated
CC region (NTR) and a single polypeptide coding region. The JEV genomic RNA,
CC JEV cDNA and reagents are useful in developing vaccines for and in
CC diagnosing and treating Japanese encephalitis. The present sequence is a
CC sequence of the invention.
XX
SQ Sequence 18563 BP; 4944 A; 4211 C; 4929 G; 4479 T; 0 U; 0 Other;

Query Match 84.8%; Score 17.8; DB 12; Length 18563;
Best Local Similarity 90.5%; Pred. No. 51;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCGGGCTGTCAATATGCTAAA 21
||||||| |||||||||| ||
Db 128 CCGGGCTATCAATATGCTGAA 148

Search completed: September 6, 2005, 20:39:24
Job time : 194.656 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 6, 2005, 17:45:55 ; Search time 1500.84 Seconds
(without alignments)
532.600 Million cell updates/sec

Title: US-10-729-421-34
Perfect score: 21
Sequence: 1 ccgggctgtcaatatgctaaa 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	17.8	84.8	425	7	CO951450	CO951450 UMC-pd12f
c	2	17.8	84.8	696	7	CN153647	CN153647 940784 MA
	3	17.8	84.8	696	7	CN155761	CN155761 943088 MA
	4	17.8	84.8	998	9	AG570358	AG570358 Mus muscu
	5	17.4	82.9	429	7	CK699541	CK699541 ZF101-P00
c	6	17.4	82.9	693	9	CE434484	CE434484 tigr-gss-
	7	17	81.0	1009	9	CNS00HWB	AL073822 Drosophil
c	8	16.8	80.0	107	9	CC888327	CC888327 SALK_1516
	9	16.8	80.0	354	7	CK091023	CK091023 F039P30.3
c	10	16.8	80.0	367	7	CK101326	CK101326 F039P30.5
	11	16.8	80.0	407	4	BG037611	BG037611 dc53e09.y
c	12	16.8	80.0	501	7	CN489600	CN489600 Mdfw20191
c	13	16.8	80.0	516	8	AZ981379	AZ981379 2M0258I13
c	14	16.8	80.0	562	5	BU815985	BU815985 N058F04 P
c	15	16.8	80.0	620	6	CA821839	CA821839 RSH08C06
	16	16.8	80.0	623	8	AZ832056	AZ832056 2M0112L09
c	17	16.8	80.0	631	7	CK317796	CK317796 B9P01h01
c	18	16.8	80.0	635	5	BU863468	BU863468 S028D11 P
	19	16.8	80.0	751	7	CV257160	CV257160 WS0245.B2
c	20	16.8	80.0	790	9	CR268537	CR268537 Reverse s
	21	16.8	80.0	900	9	CR016265	CR016265 Forward s
c	22	16.8	80.0	986	5	BQ220271	BQ220271 AGENCOURT
	23	16.4	78.1	298	9	CE199293	CE199293 tigr-gss-
c	24	16.4	78.1	363	1	AL926049	AL926049 AL926049
c	25	16.4	78.1	502	1	AI512957	AI512957 LD45125.5
	26	16.4	78.1	522	5	BX723363	BX723363 BX723363
c	27	16.4	78.1	604	8	BH039924	BH039924 RPCI-24-2
	28	16.4	78.1	686	7	CO125121	CO125121 GR_Eb08G
	29	16.4	78.1	704	9	CC870522	CC870522 NDL.123C7
c	30	16.4	78.1	815	5	BU381267	BU381267 603861468
	31	16.4	78.1	851	5	BX723938	BX723938 BX723938
c	32	16.2	77.1	223	4	BI320339	BI320339 saf21f09.
	33	16.2	77.1	403	7	CF190539	CF190539 k8k07j2.f
c	34	16.2	77.1	405	7	CN496900	CN496900 Mdfw2023o
	35	16.2	77.1	432	7	CO285402	CO285402 EK168918.
	36	16.2	77.1	503	1	AA918584	AA918584 ol53a09.s
c	37	16.2	77.1	537	8	AZ713066	AZ713066 RPCI-24-6
	38	16.2	77.1	543	6	BY566461	BY566461 BY566461
	39	16.2	77.1	551	8	AQ425764	AQ425764 CITBI-E1-
c	40	16.2	77.1	560	9	CE166745	CE166745 tigr-gss-
	41	16.2	77.1	570	5	BX711144	BX711144 BX711144
c	42	16.2	77.1	573	7	CN995009	CN995009 Mdfw2050m
c	43	16.2	77.1	577	2	BF260707	BF260707 HVSMEf002
c	44	16.2	77.1	578	4	BG000437	BG000437 MR3-GN015
c	45	16.2	77.1	592	4	BM440929	BM440929 EBro08_SQ
	46	16.2	77.1	601	2	BF614355	BF614355 de05e10.y
	47	16.2	77.1	618	2	BF676682	BF676682 602086323
	48	16.2	77.1	621	6	CA694474	CA694474 wlmk4.pk0
c	49	16.2	77.1	637	7	CN910418	CN910418 030128ABL
	50	16.2	77.1	651	4	BJ589589	BJ589589 BJ589589
c	51	16.2	77.1	652	7	CF365321	CF365321 835940 MA
	52	16.2	77.1	659	4	BJ591475	BJ591475 BJ591475
	53	16.2	77.1	698	4	BJ589606	BJ589606 BJ589606
	54	16.2	77.1	721	9	AG417300	AG417300 Mus muscu
c	55	16.2	77.1	732	9	AG558847	AG558847 Mus muscu
c	56	16.2	77.1	743	9	AG548062	AG548062 Mus muscu
c	57	16.2	77.1	744	9	AG566555	AG566555 Mus muscu
	58	16.2	77.1	756	4	BJ594699	BJ594699 BJ594699
	59	16.2	77.1	757	7	CO802419	CO802419 AGENCOURT
	60	16.2	77.1	800	9	BX997985	BX997985 Reverse s
c	61	16.2	77.1	824	8	AZ187754	AZ187754 SP_1009_B
	62	16.2	77.1	859	8	BZ701219	BZ701219 PUCCO20TD
	63	16.2	77.1	860	8	BZ704489	BZ704489 PUCCN20TD
c	64	16.2	77.1	863	7	CF374332	CF374332 AGENCOURT
	65	16.2	77.1	886	5	BX851663	BX851663 BX851663
c	66	16.2	77.1	932	9	CL240184	CL240184 ZMMBBb058
	67	16.2	77.1	936	5	BQ688620	BQ688620 AGENCOURT
c	68	16.2	77.1	939	9	CNS04C0N	AL283856 Tetraodon
	69	16.2	77.1	942	6	CD245530	CD245530 AGENCOURT
	70	16.2	77.1	953	9	CG094980	CG094980 PUKBV26TD
c	71	16.2	77.1	1009	9	CG094979	CG094979 PUKBV26TB
c	72	16.2	77.1	1016	9	CL997901	CL997901 ZMMBHf001
	73	16.2	77.1	1049	9	CL391702	CL391702 ZMMBBb019
	74	16.2	77.1	1058	8	CC221736	CC221736 CH261-92I
	75	16.2	77.1	1062	8	CC221316	CC221316 CH261-83N
c	76	16	76.2	391	4	BI028233	BI028233 CM4-MT028
	77	16	76.2	402	6	CB218410	CB218410 NISC_nb08
	78	16	76.2	535	8	AQ226974	AQ226974 HS_2013_A
c	79	16	76.2	588	9	CE471568	CE471568 tigr-gss-
c	80	16	76.2	657	8	BH352121	BH352121 CH230-220
	81	16	76.2	774	7	CK448276	CK448276 pncs900aD
c	82	16	76.2	929	9	CNS04IGB	AL292196 Tetraodon
c	83	16	76.2	959	9	CL477040	CL477040 SAIL_266_
	84	16	76.2	1050	9	CL046976	CL046976 CH216-65F
	85	15.8	75.2	220	6	CD112634	CD112634 ME1-0024T
c	86	15.8	75.2	234	7	R40815	R40815 yf82b01.s1
c	87	15.8	75.2	246	7	R61488	R61488 yh15g09.s1
	88	15.8	75.2	268	5	BP751208	BP751208 BP751208
c	89	15.8	75.2	270	1	AV210948	AV210948 AV210948
c	90	15.8	75.2	273	7	H09942	H09942 ym01c11.s1
	91	15.8	75.2	286	1	AA958331	AA958331 MAAD0514.
	92	15.8	75.2	294	1	AV228707	AV228707 AV228707
	93	15.8	75.2	295	8	CC072824	CC072824 CSU-K33r.
c	94	15.8	75.2	306	4	BI126863	BI126863 I081P66P
c	95	15.8	75.2	328	4	BI126556	BI126556 I077P15P
	96	15.8	75.2	340	8	AZ783049	AZ783049 2M0024J20
c	97	15.8	75.2	350	1	AU233672	AU233672 AU233672

98	15.8	75.2	355	1	AU278073	AU278073	AU278073
c 99	15.8	75.2	388	9	CE627437	CE627437	tigr-gss-
c 100	15.8	75.2	430	8	AQ437513	AQ437513	HS_5133_B

ALIGNMENTS

RESULT 1
CO951450
LOCUS CO951450 425 bp mRNA linear EST 09-SEP-2004
DEFINITION UMC-pd12fol-008-a09 Day 12 ovarian follicle pd12fol Sus scrofa cDNA 3', mRNA sequence.
ACCESSION CO951450
VERSION CO951450.1 GI:51328453
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 425)
AUTHORS Jiang,H., Whitworth,K.M., Bivens,N.J., Ries,J.E., Woods,R.J., Forrester,L.J., Springer,G.K., Mathialagan,N., Agca,C., Prather,R.S. and Lucy,M.C.
TITLE Large-scale Generation and Analysis of Expressed Sequence Tags from Porcine Ovary
JOURNAL Unpublished (2004)
COMMENT Contact: DNA Core Facility (Swine Project)
Animal Science - RS Prather
University of Missouri-Columbia
M616 Medical Sciences Bldg., Columbia, MO 65212, USA
Tel: (573)882-0428
Fax: (573)884-5552
Email: porcine@rnet.missouri.edu
POLYA=No.
FEATURES Location/Qualifiers
source 1. .425
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/dev_stage="Day 12 ovarian follicle"
/clone_lib="pd12fol"
/note="Vector: T3T7PAC; Funding: A grant from the Monsanto Company to the University of Missouri. Genetic Source: Ovarian tissue (whole ovary, dissected follicles, or corpora lutea) was collected from crossbred pigs (Sus scrofa domestica), frozen in liquid nitrogen shortly after collection, and stored at -80 degrees Celsius until RNA extraction. The tissue from several individual pigs was pooled for the purpose of RNA extraction. The specific tissues collected were fetal whole ovary; neonatal whole ovary; prepubutal whole ovary; 2, 4, 6 and 8 mm growing follicles; Day 0 follicles; Day 5 small antral follicles and corpora lutea; Day 12 corpora lutea and Day 12 follicles. More information regarding the methods can be found at:
http://genome.rnet.missouri.edu/Swine/Methods.html.
Library Construction (Standard Protocol): All procedures discussed in this section have been described in detail elsewhere (Soares et al., 1994; Bonaldo et al., 1996; Jiang et al., 2001). Total cellular RNA from each sample was isolated by using STAT-60 reagent (Tel-Test, Friendswood, TX) and the poly(A)+ RNA was obtained by two rounds of purification with the Oligotex mRNA isolation kit (Qiagen) according to the manufacturer's instructions. The libraries were constructed essentially as described by the manufacturer's instructions provided with the SuperScript Plasmid System (Invitrogen, cat. no. 18248-013). Briefly, 1mg of poly(A)+ RNA will be annealed at 37 degrees Celsius with 10mg of NotI-tag-dT18 oligonucleotide (GCTGCTCGCGGCCGC-tag-T18) and reverse transcribed at 37 degrees Celsius with SuperScript II (Invitrogen) reverse transcriptase (Jiang et al., 2001). The 'tag' represents a tissue/stage-specific ten-base sequence identifier (http://genome.uiowa.edu/pubsoft/software.html) present in the oligonucleotide used to prime first-strand synthesis. Second strand synthesis was performed with T4 DNA polymerase in the presence of DNA ligase and RNase H. After second strand synthesis, the double-stranded cDNAs was ligated to SalI adapters (Invitrogen-Life Technologies) and digested with NotI. The cDNAs will be size selected by passage through cDNA size fractionation columns (Invitrogen-Life technologies). The cDNAs derived from each developmental stage of a particular tissue were mixed on an equimolar basis and ligated directionally into the NotI and SalI sites of the pSPORT1 vector (Invitrogen). After ligation of the inserts, the plasmids will be electroporated into DH10B bacteria. Preliminary Library Characterization: Randomly chosen clones from each library were analyzed by restriction digestion to determine average insert size (96 clones) and by sequencing (~4 96-well plates) to confirm library quality [e.g. the presence of short polyA+ tails, genomic DNA contamination (must be <1%), ribosomal RNA clones (must be <1%), etc.] and to provide a sequence database representing the predominant clones in each library. The clones were sequenced at the University of Missouri-Columbia DNA Core Facility. Bioinformatics work was performed by GK Springer's bioinformatics group (WG Spollen, JE Ries, A Guillen, AA Khambati, RV Patel, CM Topinka, SB Bhuiyan) in Computer Science and Health Management and Informatics Departments at the University of Missouri-Columbia. Clone Requests: Requests for clones should be made to the Director of the University of Missouri DNA Core facility at: porcine@rnet.missouri.edu. Citations: 1. Bonaldo MF, Lennon G, Soares MB. Normalization and Subtraction: Two approaches to facilitate gene discovery. Genome Res, 1996; 6:791-806. 2. Jiang H, Bivens NJ, Ries JE, Whitworth KM, Green JA, Forrester LJ, Springer GK, Didion BA, Mathialagan N, Prather RS, Lucy MC (2001) Constructing cDNA libraries with fewer clones that contain long poly(dA) tails. Biotechniques 31:38-42. 3. Soares MB, MF Banaldo, P Jelene, L Su, L Lawton, A Efstrantiadis. 1994. Construction and characterization of a normalized cDNA library. Proc Natl Acad Sci, 91:9228-9232.
TAG_TISSUE=Day 12 ovarian follicle
TAG_SEQ=Not found"
ORIGIN

Query Match 84.8%; Score 17.8; DB 7; Length 425;
Best Local Similarity 90.5%; Pred. No. 2.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 CCGGGCTGTCAATATGCTAAA 21
              ||||||||  | ||||||||||
Db         49 CCGGGCTGGCGATATGCTAAA 69
```

RESULT 2
CN153647/c
LOCUS CN153647 696 bp mRNA linear EST 02-APR-2004
DEFINITION 940784 MARC 4PIG Sus scrofa cDNA 3', mRNA sequence.
ACCESSION CN153647
VERSION CN153647.1 GI:46168077
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 696)
AUTHORS Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Wise,T.A., Nonneman,D.J., Wray,J.E. and Keele,J.W.
TITLE Porcine EST collection using a normalized library constructed from embryos representing early developmental stages

```
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called with phred v0.020425.c and
            trimmed with the aid of the trim_alt option. Vector identified with
            cross_match v0.990329.
            Plate: TMW8048  row: D  column: 8
            Seq primer: TAGAAGGCACAGTCGAGG.
FEATURES             Location/Qualifiers
     source          1. .696
                     /organism="Sus scrofa"
                     /mol_type="mRNA"
                     /db_xref="taxon:9823"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 4PIG"
                     /note="Vector: pcDNA3.1; Site_1: EcoRI; Site_2: NotI;
                     Library made with combined RNA from day-10, day-13,
                     day-15, day-25, and day-30 whole embryos."
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 7;  Length 696;
  Best Local Similarity   90.5%;  Pred. No. 2.3e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCTAAA 21
              ||||||||  | ||||||||||
Db        488 CCGGGCTGGCGATATGCTAAA 468


RESULT 3
CN155761
LOCUS       CN155761                696 bp    mRNA    linear   EST 02-APR-2004
DEFINITION  943088 MARC 4PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   CN155761
VERSION     CN155761.1  GI:46170191
KEYWORDS    EST.
SOURCE      Sus scrofa (pig)
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 696)
  AUTHORS   Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Wise,T.A.,
            Nonneman,D.J., Wray,J.E. and Keele,J.W.
  TITLE     Porcine EST collection using a normalized library constructed from
            embryos representing early developmental stages
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called with phred v0.020425.c and
            trimmed with the aid of the trim_alt option. Vector identified with
            cross_match v0.990329.
            Plate: TMW8048  row: D  column: 8
            Seq primer: GTAATACGACTCACTATAGGG.
FEATURES             Location/Qualifiers
     source          1. .696
                     /organism="Sus scrofa"
                     /mol_type="mRNA"
                     /db_xref="taxon:9823"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 4PIG"
                     /note="Vector: pcDNA3.1; Site_1: EcoRI; Site_2: NotI;
                     Library made with combined RNA from day-10, day-13,
                     day-15, day-25, and day-30 whole embryos."
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 7;  Length 696;
  Best Local Similarity   90.5%;  Pred. No. 2.3e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCTAAA 21
              ||||||||  | ||||||||||
Db        209 CCGGGCTGGCGATATGCTAAA 229


RESULT 4
AG570358
LOCUS       AG570358                998 bp    DNA     linear   GSS 05-JUN-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-492K10.TJ, genomic survey
            sequence.
ACCESSION   AG570358
VERSION     AG570358.1  GI:48331078
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     BAC end Sequences of Library MSMg01
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 998)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .998
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-492K10.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 9;  Length 998;
  Best Local Similarity   90.5%;  Pred. No. 2.4e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCTAAA 21
              |||||||||||||  | |||||
Db        193 CCGGGCTGTCAATCTCCTAAA 213


RESULT 5
CK699541
LOCUS       CK699541                429 bp    mRNA    linear   EST 30-MAR-2004
DEFINITION  ZF101-P00082-DEPE-F_G24 GISZF001_ra Danio rerio cDNA clone
```

```
            IMAGE:7165298 5', mRNA sequence.
ACCESSION   CK699541
VERSION     CK699541.1  GI:42451877
KEYWORDS    EST.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 429)
  AUTHORS   Wei,C., Mathavan,S., Thoreau,H., Lim,L., Lee,C. and Ruan,Y.
  TITLE     Genome Institute of Singapore, Zebrafish Gene Collection
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Ruan Y
            Cloning and Sequencing
            Genome Institute of Singapore
            60 Biopolis Street, #02-01, Genome, Singapore 138672
            Tel: +65 6478 8073
            Fax: +65 6478 9059
            Email: ruanyj@gis.a-star.edu.sg
            GIS Clone ID: ZF101-P00082-BR2_G24
            PCR PRimers
            FORWARD: M13
            BACKWARD: M13
            Plate: ZF101-P00082-BR2  row: G  column: 24
            Seq primer: CCGCATAACTTGTATAGCA
            High quality sequence stop: 429.
FEATURES             Location/Qualifiers
     source          1. .429
                     /organism="Danio rerio"
                     /mol_type="mRNA"
                     /strain="Singapore local strain"
                     /db_xref="taxon:7955"
                     /clone="IMAGE:7165298"
                     /tissue_type="Embryo"
                     /dev_stage="7 Different embryonic Stages(From just
                     fertilized Embryos to 72 hours just hatched baby fish)"
                     /lab_host="DH10B"
                     /clone_lib="GISZF001_ra"
                     /note="Vector: pDNR-LIB; Site_1: Sfi A (GGCCATTACGGCC);
                     Site_2: Sfi B (GGCCGAGGCGGCC); Priming method: Sfi-(dT)30
                     Primed ; Priming sequence:
                     5.ATTCTAGAGGCCGAGGCGGCCGACATG(T)30VN ; Directionally
                     cloned,  5' cloning site: Sfi A site GGCCATTACGGCC ;  5'
                     linker/adaptor sequence: 5.AAGCAGTGGTATCAACGCAGAGTGGCC ;
                     3' cloning site: Sfi B site GGCCGAGGCGGCC ;  3'
                     linker/adaptor sequence: same as the priming sequence ;
                     Average insert size: 2kb ; For PCR insert analysis: Use
                     M13 Forward and reverse primers ; Library Amplified ;
                     Recombinants (inserts): 98% ; Library complexity: 5x106 ;
                     Full-length construction (method): SMART, a Clontech
                     method The pooled tissue RNA was collected and used to
                     construct full length enriched cDNA library and also
                     served as template to synthesize complex first strand cDNA
                     probe. Two high density colony arrays were made from over
                     110K cDNA clones and hybridized with the probes. Low
                     intensity clones were selected as they represented rare
                     expressed clones. The hybridization intensities for all
                     clones span from 0 to 1.8 million counts and the low
                     abundant class ranged from 0 to 13,000."
ORIGIN

  Query Match             82.9%;  Score 17.4;  DB 7;  Length 429;
  Best Local Similarity   94.7%;  Pred. No. 3.6e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 GGGCTGTCAATATGCTAAA 21
              |||||||||| ||||||||
Db          3 GGGCTGTCAAGATGCTAAA 21


RESULT 6
CE434484/c
```

```
LOCUS       CE434484                693 bp    DNA     linear   GSS 27-SEP-2003
DEFINITION  tigr-gss-dog-17000363221477 Dog Library Canis familiaris genomic,
            genomic survey sequence.
ACCESSION   CE434484
VERSION     CE434484.1  GI:36711024
KEYWORDS    GSS.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1  (bases 1 to 693)
  AUTHORS   Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
  TITLE     The dog genome: survey sequencing and comparative analysis
  JOURNAL   Science 301 (5641), 1898-1903 (2003)
  MEDLINE   22875432
   PUBMED   14512627
COMMENT     Contact: Kirkness EF
            The Institute for Genomic Research
            Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
            Rockville, MD 20850, USA
            Tel: 301-838-0200
            Fax: 301-838-0208
            Email: ekirknes@tigr.org
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .693
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="Standard Poodle"
                     /db_xref="taxon:9615"
                     /clone_lib="Dog Library"
                     /note="Site_1: BstXI; Libraries were prepared from
                     peripheral blood"
ORIGIN

  Query Match             82.9%;  Score 17.4;  DB 9;  Length 693;
  Best Local Similarity   94.7%;  Pred. No. 3.8e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 GGGCTGTCAATATGCTAAA 21
              |||||||||| ||||||||
Db        326 GGGCTGTCATTATGCTAAA 308


RESULT 7
CNS00HWB
LOCUS       CNS00HWB                1009 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR35D08 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL073822
VERSION     AL073822.1  GI:4953796
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1009)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
```

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```
FEATURES             Location/Qualifiers
     source          1. .1009
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR35D08"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match           81.0%;  Score 17;  DB 9;  Length 1009;
  Best Local Similarity  85.7%;  Pred. No. 6.4e+02;
  Matches   18;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCTAAA 21
              |||| |||||||||| |:|||
Db        133 CCGGCCTGTCAATATACHAAA 153


RESULT 8
CC888327/c
LOCUS       CC888327                107 bp    DNA     linear   GSS 31-JUL-2003
DEFINITION  SALK_151698.23.35.x Arabidopsis thaliana TDNA insertion lines
            Arabidopsis thaliana genomic clone SALK_151698.23.35.x, genomic
            survey sequence.
ACCESSION   CC888327
VERSION     CC888327.1  GI:33364847
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 107)
  AUTHORS   Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
            Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
            Shinn,P., Zimmerman,J. and Ecker,J.R.
  TITLE     A Sequence-Indexed Library of Insertion Mutations in the
            Arabidopsis Genome
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Joseph R. Ecker
            Salk Institute Genomic Analysis Laboratory (SIGnAL)
            The Salk Institute for Biological Studies
            10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
            Tel: 858 453 4100 x1752
            Fax: 858 558 6379
            Email: ecker@salk.edu
            This is single pass sequence recovered from the left border of
            TDNA.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .107
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /ecotype="Col-0"
                     /db_xref="taxon:3702"
                     /clone="SALK_151698.23.35.x"
                     /clone_lib="Arabidopsis thaliana TDNA insertion lines"
                     /note="PCR was performed on Arabidopsis thaliana lines
                     each of which contains  one or more TDNA insertion
                     elements. The resultant fragment for each  line was
                     directly sequenced to determine the genomic sequence at
                     the  site of insertion. Details of the protocols used can
                     be found at  http://signal.salk.edu/tdna_protocols.html"
ORIGIN

  Query Match           80.0%;  Score 16.8;  DB 9;  Length 107;
  Best Local Similarity  90.0%;  Pred. No. 6.4e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCTAA 20
              |||||||||||||| || ||
Db         79 CCGGGCTGTCAATAAGCCAA 60


RESULT 9
CK091023
LOCUS       CK091023                354 bp    mRNA    linear   EST 01-DEC-2003
DEFINITION  F039P30.3pR Populus flower cDNA library Populus balsamifera subsp.
            trichocarpa cDNA clone F039P30 3', mRNA sequence.
ACCESSION   CK091023
VERSION     CK091023.1  GI:38575348
KEYWORDS    EST.
SOURCE      Populus balsamifera subsp. trichocarpa (Populus trichocarpa)
  ORGANISM  Populus balsamifera subsp. trichocarpa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Malpighiales; Salicaceae; Saliceae; Populus.
REFERENCE   1  (bases 1 to 354)
  AUTHORS   Sterky,F., Bhalerao,R.R., Unneberg,P., Segerman,B., Nilsson,P.,
            Brunner,A.M., Campaa,L., Jonsson-Lindvall,J., Tandre,K.,
            Strauss,S.H., Sundberg,B., Gustafsson,P., Uhlen,M., Bhalerao,R.P.,
            Nilsson,O., Sandberg,G., Karlsson,J., Lundeberg,J. and Jansson,S.
  TITLE     A Populus EST resource for functional genomics
  JOURNAL   Unpublished (2003)
COMMENT     Other_ESTs: F039P30Y, F039P30.5pR
            Contact: Bo Segerman
            Umea Plant Science Center, Department of Plant Physiology
            Umea University
            901 87 Umea, Sweden
            Tel: +46 90 786 5279
            Fax: +46 90 786 6676
            Email: bo.segerman@plantphys.umu.se.
FEATURES             Location/Qualifiers
     source          1. .354
                     /organism="Populus balsamifera subsp. trichocarpa"
                     /mol_type="mRNA"
                     /sub_species="trichocarpa"
                     /db_xref="taxon:3694"
                     /clone="F039P30"
                     /tissue_type="floral buds"
                     /clone_lib="Populus flower cDNA library"
                     /note="Organ: flower"
ORIGIN

  Query Match           80.0%;  Score 16.8;  DB 7;  Length 354;
  Best Local Similarity  90.0%;  Pred. No. 7.3e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 CGGGCTGTCAATATGCTAAA 21
              |||||| |||| |||||||||
Db        213 CGGGCTTTCAAAATGCTAAA 232


RESULT 10
CK101326/c
LOCUS       CK101326                367 bp    mRNA    linear   EST 01-DEC-2003
DEFINITION  F039P30.5pR Populus flower cDNA library Populus balsamifera subsp.
            trichocarpa cDNA clone F039P30 5', mRNA sequence.
ACCESSION   CK101326
VERSION     CK101326.1  GI:38585651
KEYWORDS    EST.
SOURCE      Populus balsamifera subsp. trichocarpa (Populus trichocarpa)
  ORGANISM  Populus balsamifera subsp. trichocarpa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Malpighiales; Salicaceae; Saliceae; Populus.
REFERENCE   1  (bases 1 to 367)
```

```
  AUTHORS   Sterky,F., Bhalerao,R.R., Unneberg,P., Segerman,B., Nilsson,P.,
            Brunner,A.M., Campaa,L., Jonsson-Lindvall,J., Tandre,K.,
            Strauss,S.H., Sundberg,B., Gustafsson,P., Uhlen,M., Bhalerao,R.P.,
            Nilsson,O., Sandberg,G., Karlsson,J., Lundeberg,J. and Jansson,S.
  TITLE     A Populus EST resource for functional genomics
  JOURNAL   Unpublished (2003)
COMMENT     Other_ESTs: F039P30Y, F039P30.3pR
            Contact: Bo Segerman
            Umea Plant Science Center, Department of Plant Physiology
            Umea University
            901 87 Umea, Sweden
            Tel: +46 90 786 5279
            Fax: +46 90 786 6676
            Email: bo.segerman@plantphys.umu.se.
FEATURES             Location/Qualifiers
     source          1. .367
                     /organism="Populus balsamifera subsp. trichocarpa"
                     /mol_type="mRNA"
                     /sub_species="trichocarpa"
                     /db_xref="taxon:3694"
                     /clone="F039P30"
                     /tissue_type="floral buds"
                     /clone_lib="Populus flower cDNA library"
                     /note="Organ: flower"
ORIGIN

  Query Match             80.0%;  Score 16.8;  DB 7;  Length 367;
  Best Local Similarity   90.0%;  Pred. No. 7.4e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 CGGGCTGTCAATATGCTAAA 21
              |||||| |||| ||||||||
Db        318 CGGGCTTTCAAGATGCTAAA 299


RESULT 11
BG037611
LOCUS       BG037611                407 bp    mRNA    linear   EST 24-JAN-2001
DEFINITION  dc53e09.y1 NICHD_XGC_Emb3 Xenopus laevis cDNA clone IMAGE:3400816
            5', mRNA sequence.
ACCESSION   BG037611
VERSION     BG037611.1  GI:12480196
KEYWORDS    EST.
SOURCE      Xenopus laevis (African clawed frog)
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Xenopus.
REFERENCE   1  (bases 1 to 407)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Other_ESTs: dc53e09.x1
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Martha Rebbert, Steven L. Klein, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: Xenopus clones from this library are available
            through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
            Seq primer: -40RP from Gibco
            High quality sequence stop: 403.
FEATURES             Location/Qualifiers
     source          1. .407
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:3400816"
                     /tissue_type="embryo (stages 24-25)"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NICHD_XGC_Emb3"
                     /note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI;
                     Cloned unidirectionally. Primer: Oligo dT. Average insert
                     size 1.7 kb. Constructed by Life Technologies. Note: This
                     is a Xenopus Gene Collection (XGC) library."
ORIGIN

  Query Match             80.0%;  Score 16.8;  DB 4;  Length 407;
  Best Local Similarity   85.7%;  Pred. No. 7.4e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCTAAA 21
              | ||| ||| ||||||||||
Db        364 CAGGGTTGTNAATATGCTAAA 384


RESULT 12
CN489600/c
LOCUS       CN489600                501 bp    mRNA    linear   EST 24-MAY-2004
DEFINITION  Mdfw2019108.y1 Mdfw Malus x domestica cDNA clone Mdfw2019108 5'
            similar to TR:Q9SZR6 Q9SZR6 HYPOTHETICAL 31.9 KD PROTEIN. ;, mRNA
            sequence.
ACCESSION   CN489600
VERSION     CN489600.1  GI:46603708
KEYWORDS    EST.
SOURCE      Malus x domestica (cultivated apple)
  ORGANISM  Malus x domestica
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Rosales; Rosaceae; Maloideae; Malus.
REFERENCE   1  (bases 1 to 501)
  AUTHORS   Korban,S., Vodkin,L., Liu,L., Gasic,K., Gonzales,O., Hernandez,A.,
            Aldwinckle,H., Malnoy,M., Carroll,N., Goldsbrough,P., Orvis,K.,
            Clifton,S., Pape,D., Marra,M., Hillier,L., Martin,J., Wylie,T.,
            Dante,M., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Ronko,I.,
            Tsagareishvili,R., Kennedy,S., Waterston,R. and Wilson,R.
  TITLE     Apple Functional Genomics grant - NSF 0321702
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Schuyler S. Korban
            Apple Functional Genomics grant - NSF 0321702
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library materials provided by: Schuyler S. Korban Library
            constructed by: A. Hernandez / K. Gasic Library sequenced by:
            Washington University Genome Sequencing Center
            WashU EST name: aaf76f04.y1
            High quality sequence stop: 425.
FEATURES             Location/Qualifiers
     source          1. .501
                     /organism="Malus x domestica"
                     /mol_type="mRNA"
                     /db_xref="taxon:3750"
                     /clone="Mdfw2019108"
                     /lab_host="DH10B ampicillin resistant"
                     /clone_lib="Mdfw"
                     /note="Vector: DH10B ampicillin resistant; Site_1: NotI;
                     Site_2: EcoRII; Total RNA was extracted separately from
                     each stage (bud, balloon, open and after pollination),
                     using the 'pine tree' method. Poly(A)+mRNA was isolated
                     twice from total RNA from each stage using the Oligotex
                     Direct mRNA kit (Qiagen). mRNA was reverse transcribed
                     into double stranded cDNA using a modified oligo18(dT)
                     primer with an identifying tag sequence (see table below).
                     cDNAs from different stages were pooled in equal amounts
                     before adaptor ligation. Tag identification when
                     sequencing from 5' end: Stage 1 (bud) insert 18(A)TCGGA;
                     Stage 2 (balloon) insert 18(A)TGCGA; Stage 3 (open) insert
                     18(A)TGCGT; Stage 4 (after pollination) insert 18(A)TCGGT.
                     Tag identification when sequencing from 3' end: Stage 1
                     (bud) TCCGA18(T) insert; Stage 2 (balloon) TCGCA18(T)
                     insert; Stage 3 (open) ACGCA18(T) insert; Stage 4 (after
```

pollination) ACCGA18(T) insert. Double stranded cDNAs were size selected (more than 450 bp), adaptored with EcoRI adapters at both ends and then digested with NotI. The cDNAs were then directionally cloned into EcoR1-NotI digested pBS II SK(+) phagemid vector(Stratagene). Identification of adaptors and tags in 5'-end sequenced clones: <Vector>. . .TAAGCTT<End Vector><Start EcoR1_adaptor>GATATCGAATTCCATTGTGTTGGG <End EcoR1_adaptor><Start Insert>. . . .AAAAAAAAAAAAAAAAAA<End Insert> <Start Tag>TGCGA<End Tag><Start NotI_site/Vector>GCGGCCGCCACCGCGG. . . The total number of white colony forming units (cfu) in the primary library before amplification was 1.1x10^6 cfu (colony forming units). The background of empty clones was less than 1%. Inserts ranged from 0.5kb to 3 kb, as determined by PCR. Purified plasmid DNA from the primary library was converted to single-stranded circles and used as a template for PCR amplification using the T7 and T3 priming sites flanking the cloned cDNA inserts. The purified PCR products, representing the entire cloned cDNA population, were used as a driver for normalization. Hybridization between the single-stranded library and the PCR products was carried out for 44 hours at 30C. Unhybridized single-stranded DNA circles were separated from hybridized DNA rendered partially double-stranded and electroporated into DH10B cells to generate the normalized library. The total number of clones with insert was 9x10^6 cfu. Background of empty clones was less than 1%"

ORIGIN

Query Match 80.0%; Score 16.8; DB 7; Length 501;
Best Local Similarity 85.7%; Pred. No. 7.6e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy          1 CCGGGCTGTCAATATGCTAAA 21
              |||||||||| || ||| |||
Db        332 CCGGGCTGTCCATTTGCNAAA 312
```

RESULT 13
AZ981379/c
LOCUS AZ981379 516 bp DNA linear GSS 27-APR-2001
DEFINITION 2M0258I13R Mouse 10kb plasmid UUGC2M library Mus musculus genomic clone UUGC2M0258I13 R, genomic survey sequence.
ACCESSION AZ981379
VERSION AZ981379.1 GI:13852606
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 516)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0258 row: I column: 13
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 516.
FEATURES Location/Qualifiers
source 1. .516
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0258I13"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUGC2M library"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN

Query Match 80.0%; Score 16.8; DB 8; Length 516;
Best Local Similarity 90.0%; Pred. No. 7.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          2 CGGGCTGTCAATATGCTAAA 21
              | ||||||||||||||||  |
Db         30 CTGGCTGTCAATATGCTACA 11
```

RESULT 14
BU815985/c
LOCUS BU815985 562 bp mRNA linear EST 15-OCT-2002
DEFINITION N058F04 Populus bark cDNA library Populus tremula x Populus tremuloides cDNA 5 prime, mRNA sequence.
ACCESSION BU815985
VERSION BU815985.1 GI:23975607
KEYWORDS EST.
SOURCE Populus tremula x Populus tremuloides
ORGANISM Populus tremula x Populus tremuloides
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Malpighiales; Salicaceae; Saliceae; Populus.
REFERENCE 1 (bases 1 to 562)
AUTHORS Unneberg,P., Bhalerao,R.R., Jansson,S. and Sterky,F.
TITLE The poplar tree transcriptome: Analysis of expressed sequence tags from multiple libraries
JOURNAL Unpublished (2002)
COMMENT Contact: BHALERAO RUPALI R.
Umea Plant Science Center
Department of Plant Physiology
University of Umea, 901 87 Umea, Sweden
Tel: +46 90 786 5279
Fax: +46 90 786 6676
Email: rupali.bhalerao@plantphys.umu.se.
FEATURES Location/Qualifiers
source 1. .562
/organism="Populus tremula x Populus tremuloides"
/mol_type="mRNA"
/db_xref="taxon:47664"
/tissue_type="bark"
/clone_lib="Populus bark cDNA library"
ORIGIN

```
  Query Match            80.0%;  Score 16.8;  DB 5;  Length 562;
  Best Local Similarity   90.0%;  Pred. No. 7.7e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 CGGGCTGTCAATATGCTAAA 21
              |||||| |||| ||||||||
Db        293 CGGGCTTTCAAGATGCTAAA 274


RESULT 15
CA821839/c
LOCUS       CA821839                620 bp    mRNA    linear   EST 28-FEB-2003
DEFINITION  RSH08C06 two-month-old roots from clone 'Beaupre' grown for 19 days
            under restricted irrigation Populus balsamifera subsp. trichocarpa
            x Populus deltoides cDNA 5', mRNA sequence.
ACCESSION   CA821839
VERSION     CA821839.1  GI:28605388
KEYWORDS    EST.
SOURCE      Populus balsamifera subsp. trichocarpa x Populus deltoides
  ORGANISM  Populus balsamifera subsp. trichocarpa x Populus deltoides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Malpighiales; Salicaceae; Saliceae; Populus.
REFERENCE   1  (bases 1 to 620)
  AUTHORS   Kohler,A., Delaruelle,C., Martin,D., Encelot,N. and Martin,F.
  TITLE     The poplar root transcriptome: analysis of 7000 expressed sequence
            tags
  JOURNAL   FEBS Lett. 542 (1-3), 37-41 (2003)
COMMENT     Contact: Martin FM
            Equipe de Microbiologie Forestiere
            Institut National de la Recherche Agronomique
            Centre INRA de Nancy, 54280 Champenoux, France
            Tel: +33 383 39 40 80
            Fax: +33 383 39 40 69
            Email: fmartin@nancy.inra.fr
            Insert Length: 620   Std Error: 0.00
            Seq primer: Fornat 5' AAGCGCGCCATTGTGTTGGTACCC.
FEATURES             Location/Qualifiers
     source          1. .620
                     /organism="Populus balsamifera subsp. trichocarpa x
                     Populus deltoides"
                     /mol_type="mRNA"
                     /cultivar="'Beaupre'"
                     /db_xref="taxon:3695"
                     /dev_stage="two-month-old"
                     /clone_lib="two-month-old roots from clone 'Beaupre' grown
                     for 19 days under restricted irrigation"
                     /note="Organ: root; Vector: pTriplEx2; cDNA library of
                     roots from two-month-old Populus trichocarpa Torr.&  Gray
                     x deltoides Bartr. Ex Marshall (clone 'Beaupre') grown for
                     19  days under restricted irrigation to reach 50% of the
                     transpiration  rate of fully watered plants. The cDNA
                     library was constructed from 1 ug of total RNA using the
                     SMART cDNA synthesis kit (Clontech, Palo Alto, CA, USA)
                     according to the manufacturer's instructions. The
                     resulting cDNA was packed into lambda phages using the
                     Gigapack III Gold packaging kit (Stratagene, La Jolla,
                     CA). The pTriplEx2 phagemid clones in Escherichia choli
                     were obtained by using the mass in vivo excision protocol
                     according to the manufacturer's instructions (Clontech)."
ORIGIN

  Query Match            80.0%;  Score 16.8;  DB 6;  Length 620;
  Best Local Similarity   90.0%;  Pred. No. 7.8e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 CGGGCTGTCAATATGCTAAA 21
              |||||| |||| ||||||||
Db        389 CGGGCTTTCAAGATGCTAAA 370


RESULT 16
AZ832056
```

```
LOCUS       AZ832056                623 bp    DNA     linear   GSS 20-FEB-2001
DEFINITION  2M0112L09F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0112L09 F, genomic survey sequence.
ACCESSION   AZ832056
VERSION     AZ832056.1  GI:13001964
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 623)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
            Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
            Niederhausern,A. and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0112  row: L  column: 09
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 623.
FEATURES             Location/Qualifiers
     source          1. .623
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0112L09"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
ORIGIN

  Query Match            80.0%;  Score 16.8;  DB 8;  Length 623;
  Best Local Similarity   90.0%;  Pred. No. 7.8e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 CGGGCTGTCAATATGCTAAA 21
              | ||||||||||||||| |
Db        563 CTGGCTGTCAATATGCTACA 582


RESULT 17
CK317796/c
```

```
LOCUS       CK317796                631 bp    mRNA    linear   EST 11-MAY-2004
DEFINITION  B9P01h01 Populus stem seasonal library Populus deltoides cDNA, mRNA
            sequence.
ACCESSION   CK317796
VERSION     CK317796.1  GI:47106219
KEYWORDS    EST.
SOURCE      Populus deltoides
  ORGANISM  Populus deltoides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Malpighiales; Salicaceae; Saliceae; Populus.
REFERENCE   1  (bases 1 to 631)
  AUTHORS   Park,S. and Han,K.-H.
  TITLE     Gene expression profile during seasonal growth cycle in poplar tree
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Kyung-Hwan Han
            Department of Forestry
            Michigan State University
            126 Natural Resources, East Lansing, MI 48824-1222, USA
            Tel: 517 353 4751
            Fax: 517 432 1143
            Email: hanky@msu.edu.
FEATURES             Location/Qualifiers
     source          1. .631
                     /organism="Populus deltoides"
                     /mol_type="mRNA"
                     /strain="ILL-129"
                     /db_xref="taxon:3696"
                     /tissue_type="stem"
                     /dev_stage="1 year old"
                     /clone_lib="Populus stem seasonal library"
ORIGIN

  Query Match             80.0%;  Score 16.8;  DB 7;  Length 631;
  Best Local Similarity   90.0%;  Pred. No. 7.8e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 CGGGCTGTCAATATGCTAAA 21
              |||||| |||| ||||||||
Db        355 CGGGCTTTCAAGATGCTAAA 336


RESULT 18
BU863468/c
LOCUS       BU863468                635 bp    mRNA    linear   EST 16-OCT-2002
DEFINITION  S028D11 Populus imbibed seed cDNA library Populus tremula cDNA 5
            prime, mRNA sequence.
ACCESSION   BU863468
VERSION     BU863468.1  GI:24049528
KEYWORDS    EST.
SOURCE      Populus tremula
  ORGANISM  Populus tremula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Malpighiales; Salicaceae; Saliceae; Populus.
REFERENCE   1  (bases 1 to 635)
  AUTHORS   Unneberg,P., Bhalerao,R.R., Jansson,S. and Sterky,F.
  TITLE     The poplar tree transcriptome: Analysis of expressed sequence tags
            from multiple libraries
  JOURNAL   Unpublished (2002)
COMMENT     Contact: BHALERAO RUPALI R.
            Umea Plant Science Center
            Department of Plant Physiology
            University of Umea, 901 87 Umea, Sweden
            Tel: +46 90 786 5279
            Fax: +46 90 786 6676
            Email: rupali.bhalerao@plantphys.umu.se.
FEATURES             Location/Qualifiers
     source          1. .635
                     /organism="Populus tremula"
                     /mol_type="mRNA"
                     /db_xref="taxon:113636"
                     /tissue_type="imbibed seed"
                     /clone_lib="Populus imbibed seed cDNA library"
ORIGIN

  Query Match             80.0%;  Score 16.8;  DB 5;  Length 635;
  Best Local Similarity   90.0%;  Pred. No. 7.8e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 CGGGCTGTCAATATGCTAAA 21
              |||||| |||| ||||||||
Db        340 CGGGCTTTCAAGATGCTAAA 321


RESULT 19
CV257160
LOCUS       CV257160                751 bp    mRNA    linear   EST 22-SEP-2004
DEFINITION  WS0245.B21_N14 PTxD-ICC-N-A-14 Populus balsamifera subsp.
            trichocarpa x Populus deltoides cDNA clone WS0245_N14 3', mRNA
            sequence.
ACCESSION   CV257160
VERSION     CV257160.1  GI:52510135
KEYWORDS    EST.
SOURCE      Populus balsamifera subsp. trichocarpa x Populus deltoides
  ORGANISM  Populus balsamifera subsp. trichocarpa x Populus deltoides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Malpighiales; Salicaceae; Saliceae; Populus.
REFERENCE   1  (bases 1 to 751)
  AUTHORS   Ralph,S., Cooper,D., Kolosova,N., Oddy,C., Butterfield,Y.,
            Kirkpatrick,R., Liu,J., Palmquist,D., Stott,J., Barber,S., Yang,G.,
            Babakaiff,R., Brown-John,M., Chand,S., Featherstone,R., Masson,A.,
            Mayo,M., Moran,J., Olson,T., Wong,D., Ritland,C.E., Siddiqui,A.,
            Holt,R., Jones,S., Marra,M., Ellis,B.E., Douglas,C., Ritland,K. and
            Bohlmann,J.
  TITLE     The poplar transcriptome: Analysis of expressed sequence tags from
            multiple cDNA libraries
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Joerg Bohlmann
            Genome BC forest genomics program
            University of British Columbia
            UBC Biotechnology Laboratory, 6174 University Boulevard, Rm. 237,
            Vancouver, British Columbia, Canada, V6T 1Z3
            Tel: 1-604-822-0282
            Fax: 1-604-822-6097
            Email: bohlmann@interchange.ubc.ca
            Plate: WS0245  row: N  column: 14
            High quality sequence stop: 751
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .751
                     /organism="Populus balsamifera subsp. trichocarpa x
                     Populus deltoides"
                     /mol_type="mRNA"
                     /cultivar="H11-11"
                     /db_xref="taxon:3695"
                     /clone="WS0245_N14"
                     /sex="Male"
                     /lab_host="E. coli DH10B T1 phage resistant cells"
                     /clone_lib="PTxD-ICC-N-A-14"
                     /note="Vector: pBluescript II SK (+) XR; Site_1: EcoRI (5'
                     end of cDNA); Site_2: XhoI (3' end of cDNA); Cultured
                     cells [de Sa MM et al. (1992) Plant Physiology 98:728-737]
                     were grown in media (45mL) supplemented with either 50uM
                     salicylic acid, 50uM benzothiadiazole, 50uM methyl
                     jasmonate, 20ug chitosan or 200uL of Pollacia radiosa
                     extract. Cells were harvested after a 3 hour treatment,
                     along with untreated control cells. mRNA was isolated from
                     each tissue source independently and equal quantities of
                     mRNA from each tissue were then pooled. cDNA was prepared
                     from 5 micrograms of mRNA and directionally ligated into
                     the pBluescript II SK (+) XR vector using the pBluescript
                     II XR cDNA Library Construction Kit according to
                     manufacturer's instructions with modifications
                     (Stratagene). Plasmid DNA was then transformed by
```

electroporation into DH10B cells (Invitrogen) for propagation. Normalization was applied according to published methods [Bonaldo M.F. et al. (1996) Genome Research 6(9):791] in order to reduce the abundance of highly expressed transcripts."
ORIGIN

Query Match 80.0%; Score 16.8; DB 7; Length 751;
Best Local Similarity 90.0%; Pred. No. 8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 CGGGCTGTCAATATGCTAAA 21
||||||||||||||||||||
Db 491 CGGGCTTTCAAGATGCTAAA 510

RESULT 20
CR268537/c
LOCUS CR268537 790 bp DNA linear GSS 06-JUL-2004
DEFINITION Reverse strand read from insert in 5'HPRT insertion targeting and chromosome engineering clone MHPN344a20, genomic survey sequence.
ACCESSION CR268537
VERSION CR268537.1 GI:50047390
KEYWORDS GSS; genome survey sequence; MICER.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 790)
AUTHORS Adams,D.J., Biggs,P.J., Cox,A.V., Davies,R.M., van der Weyden,L., Jonkers,J., Smith,J., Plumb,R.W., Taylor,R.G., Nishijima,I., Yu,Y., Rogers,J. and Bradley,A.
TITLE Direct Submission
JOURNAL Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. http://www.sanger.ac.uk/MICER
FEATURES Location/Qualifiers
source 1. .790
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="MHPN344a20"
/clone_lib="MHPN"
ORIGIN

Query Match 80.0%; Score 16.8; DB 9; Length 790;
Best Local Similarity 90.0%; Pred. No. 8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 CGGGCTGTCAATATGCTAAA 21
||||||||||||||||||||
Db 361 CTGGCTGTCAATATGCTACA 342

RESULT 21
CR016265
LOCUS CR016265 900 bp DNA linear GSS 05-JUL-2004
DEFINITION Forward strand read from insert in 3'HPRT insertion targeting and chromosome engineering clone MHPP8d17, genomic survey sequence.
ACCESSION CR016265
VERSION CR016265.1 GI:49749320
KEYWORDS GSS; genome survey sequence; MICER.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 900)
AUTHORS Adams,D.J., Biggs,P.J., Cox,A.V., Davies,R.M., van der Weyden,L., Jonkers,J., Smith,J., Plumb,R.W., Taylor,R.G., Nishijima,I., Yu,Y., Rogers,J. and Bradley,A.
TITLE Direct Submission
JOURNAL Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. http://www.sanger.ac.uk/MICER
FEATURES Location/Qualifiers
source 1. .900
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="MHPP8d17"
/clone_lib="MHPP"
ORIGIN

Query Match 80.0%; Score 16.8; DB 9; Length 900;
Best Local Similarity 90.0%; Pred. No. 8.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCGGGCTGTCAATATGCTAA 20
||||||||||||||||||||
Db 145 CCGGGCTGTCAAGATGCTCA 164

RESULT 22
BQ220271/c
LOCUS BQ220271 986 bp mRNA linear EST 02-MAY-2002
DEFINITION AGENCOURT_7572589 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6044520 5', mRNA sequence.
ACCESSION BQ220271
VERSION BQ220271.1 GI:20401671
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 986)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13287 row: e column: 01
High quality sequence stop: 149.
FEATURES Location/Qualifiers
source 1. .986
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6044520"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_92"
/note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."
ORIGIN

Query Match 80.0%; Score 16.8; DB 5; Length 986;
Best Local Similarity 90.0%; Pred. No. 8.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 CGGGCTGTCAATATGCTAAA 21
||||||||||||||||||||
Db 758 CAGGCTGTCGATATGCTAAA 739

RESULT 23
CE199293
LOCUS CE199293 298 bp DNA linear GSS 25-SEP-2003
DEFINITION tigr-gss-dog-17000372211318 Dog Library Canis familiaris genomic,

```
            genomic survey sequence.
ACCESSION   CE199293
VERSION     CE199293.1  GI:35354946
KEYWORDS    GSS.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1  (bases 1 to 298)
  AUTHORS   Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
  TITLE     The dog genome: survey sequencing and comparative analysis
  JOURNAL   Science 301 (5641), 1898-1903 (2003)
  MEDLINE   22875432
   PUBMED   14512627
COMMENT     Contact: Kirkness EF
            The Institute for Genomic Research
            Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
            Rockville, MD 20850, USA
            Tel: 301-838-0200
            Fax: 301-838-0208
            Email: ekirknes@tigr.org
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .298
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="Standard Poodle"
                     /db_xref="taxon:9615"
                     /clone_lib="Dog Library"
                     /note="Site_1: BstXI; Libraries were prepared from
                     peripheral blood"
ORIGIN

  Query Match             78.1%;  Score 16.4;  DB 9;  Length 298;
  Best Local Similarity   94.4%;  Pred. No. 1.2e+03;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 GGCTGTCAATATGCTAAA 21
              |||||| ||||||||||
Db        241 GGCTGTAAATATGCTAAA 258


RESULT 24
AL926049/c
LOCUS       AL926049                363 bp    mRNA    linear   EST 06-JUL-2004
DEFINITION  AL926049 PJR-Z1+Z2 Danio rerio cDNA clone 164-D08-2, mRNA sequence.
ACCESSION   AL926049
VERSION     AL926049.1  GI:23192629
KEYWORDS    EST.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 363)
  AUTHORS   Lo,J., Lee,S., Xu,M., Liu,F., Ruan,H., Eun,A., He,Y., Ma,W.,
            Wang,W., Wen,Z. and Peng,J.
  TITLE     15000 unique zebrafish EST clusters and their future use in
            microarray for profiling gene expression patterns during
            embryogenesis
  JOURNAL   Genome Res. 13 (3), 455-466 (2003)
  MEDLINE   22505427
   PUBMED   12618376
COMMENT     Contact: Peng J
            Lab of Functional Genomics
            Institute of Molecular and Cell Biology
            30 Medical Drive, Singapore, 117609, Singapore
            Email: pengjr@imcb.a-star.edu.sg
            Clone requests: info@openbiosystems.com
            Open Biosystems,
            6705 Odyssey Drive, Huntsville, AL 35806.
FEATURES             Location/Qualifiers
     source          1. .363
                     /organism="Danio rerio"
                     /mol_type="mRNA"
                     /strain="local wildtype"
                     /db_xref="taxon:7955"
                     /clone="164-D08-2"
                     /tissue_type="whole embryo or fish"
                     /dev_stage="mixed stages"
                     /clone_lib="PJR-Z1+Z2"
ORIGIN

  Query Match             78.1%;  Score 16.4;  DB 1;  Length 363;
  Best Local Similarity   94.4%;  Pred. No. 1.2e+03;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 GGCTGTCAATATGCTAAA 21
              ||||||||| ||||||||
Db        239 GGCTGTCAAGATGCTAAA 222


RESULT 25
AI512957/c
LOCUS       AI512957                502 bp    mRNA    linear   EST 19-APR-2001
DEFINITION  LD45125.5prime LD Drosophila melanogaster embryo pOT2 Drosophila
            melanogaster cDNA clone LD45125 5prime similar to D83486: Su(fu)
            FBgn0005355 PID:g1208417 SPTREMBL:Q27279, mRNA sequence.
ACCESSION   AI512957
VERSION     AI512957.1  GI:4422375
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 502)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 451  row: C  column: 1
            High quality sequence stop: 179.
FEATURES             Location/Qualifiers
     source          1. .502
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /db_xref="taxon:7227"
                     /clone="LD45125"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="XL1 Blue"
                     /clone_lib="LD Drosophila melanogaster embryo pOT2"
                     /note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2. "
ORIGIN

  Query Match             78.1%;  Score 16.4;  DB 1;  Length 502;
  Best Local Similarity   94.4%;  Pred. No. 1.2e+03;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CCGGGCTGTCAATATGCT 18
              ||||| ||||||||||||
Db        363 CCGGGTTGTCAATATGCT 346


Search completed: September  6, 2005, 21:55:48
```

Job time : 1507.84 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 6, 2005, 21:56:10 ; Search time 1636 Seconds
(without alignments)
535.133 Million cell updates/sec

Title: US-10-729-421-8
Perfect score: 23
Sequence: 1 tcatgactgcaattccggtcttt 23

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 34239544 seqs, 19032134700 residues

Word size : 10

Total number of hits satisfying chosen parameters: 61

Minimum DB seq length: 0
Maximum DB seq length: 60

Post-processing: Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	12	52.2	55	1	AA619888	vl58h07.s
2	11	47.8	29	8	CC456749	SALK_1002
3	11	47.8	33	8	AZ817376	2M0086N22
c 4	11	47.8	41	8	BH791854	SALK_0618
c 5	11	47.8	41	8	BH812487	SALK_0618
c 6	11	47.8	50	4	BG153713	nag57g06.
c 7	11	47.8	52	2	AW249907	2821659.3
c 8	11	47.8	55	7	CN564600	tag20b04.
c 9	11	47.8	57	8	BH918919	3526_1_63
10	10	43.5	19	9	AJ588850	Arabidops
11	10	43.5	25	9	AG197078	Pan trogl
12	10	43.5	34	9	BX660142	Arabidops
c 13	10	43.5	36	8	BH810737	SALK_0511
c 14	10	43.5	36	9	DME546528	Drosophil
15	10	43.5	37	1	AA972482	op42d03.s
c 16	10	43.5	37	1	AI937582	wp81b11.x
c 17	10	43.5	37	7	U44311	U44311 ENU44311 As
18	10	43.5	37	9	BX531847	Arabidops
19	10	43.5	39	1	AV833097	AV833097
c 20	10	43.5	39	5	BX568339	BX568339
21	10	43.5	39	9	TA160H03P	AL473258 T. brucei
22	10	43.5	41	9	CR397281	Arabidops
c 23	10	43.5	42	9	BX943895	Arabidops
c 24	10	43.5	44	9	AL752437	Arabidops
c 25	10	43.5	46	6	CF049474	CF049474 QCL37a04.
c 26	10	43.5	46	7	CN753215	CN753215 ApHL3LD-X
27	10	43.5	47	8	BH000511	BH000511 2M0288M20
c 28	10	43.5	48	4	BG253356	BG253356 602362952
29	10	43.5	48	9	BX945507	BX945507 Arabidops
30	10	43.5	48	9	CR356946	CR356946 Arabidops
c 31	10	43.5	48	9	CL528770	CL528770 ASV9G01.f
c 32	10	43.5	49	5	BQ100687	BQ100687 ij22c04.x
c 33	10	43.5	49	6	CB305243	CB305243 3'EST-Nfl
34	10	43.5	49	7	CO733304	CO733304 SlLT02c05
35	10	43.5	50	1	AU103081	AU103081 AU103081
36	10	43.5	50	1	AU103082	AU103082 AU103082
c 37	10	43.5	50	8	BH612727	BH612727 SALK_0331
c 38	10	43.5	52	1	AA068274	AA068274 mm53c01.r
39	10	43.5	52	2	BF632337	BF632337 NF018E03D
40	10	43.5	52	8	AZ629385	AZ629385 1M0482N16
c 41	10	43.5	52	9	BX122966	BX122966 Danio rer
c 42	10	43.5	53	4	BG524434	BG524434 42-53 Ste
c 43	10	43.5	53	7	CN870218	CN870218 001204AAO
c 44	10	43.5	53	8	BH252021	BH252021 SALK_0124
45	10	43.5	54	8	AZ576149	AZ576149 AST-T11C0

ALIGNMENTS

RESULT 1
AA619888/c
LOCUS AA619888 55 bp mRNA linear EST 09-OCT-1997
DEFINITION vl58h07.s1 Knowles Solter mouse 2 cell Mus musculus cDNA clone IMAGE:976477 5', mRNA sequence.
ACCESSION AA619888
VERSION AA619888.1 GI:2523764
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 55)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:557205.
FEATURES Location/Qualifiers
source 1. .55
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J x DBA/2J F1"
/db_xref="taxon:10090"
/clone="IMAGE:976477"
/tissue_type="embryo"
/dev_stage="2-cell"
/lab_host="DH10B"
/clone_lib="Knowles Solter mouse 2 cell"
/note="Organ: embryo; Vector: pBluescribe (modified); Site_1: MluI; Site_2: SalI; Cloned unidirectionally from mRNA prepared from 13,500 2-cell stage embryos. Primer: SalI(dT): 5'-CGGTCGACCGTCGACCGTTTTTTTTTTTTTT-3'. cDNAs were cloned into the MluI/SalI sites of a modified pBluescribe vector using commercial linkers (NEB).

Average insert size: 1.2 kb."
ORIGIN

Query Match 52.2%; Score 12; DB 1; Length 55;
Best Local Similarity 100.0%; Pred. No. 7.5e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 GACTGCAATTCC 16
||||||||||||
Db 21 GACTGCAATTCC 10

RESULT 2
CC456749
LOCUS CC456749 29 bp DNA linear GSS 30-MAY-2003
DEFINITION SALK_100255.25.15.x Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_100255.25.15.x, genomic survey sequence.
ACCESSION CC456749
VERSION CC456749.1 GI:31217770
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 29)
AUTHORS Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J. and Ecker,J.R.
TITLE A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
JOURNAL Unpublished (2001)
COMMENT Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGnAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of TDNA. This sequence lies within 300 bases of the 5' end of At2g27775.
Class: TDNA tagged.
FEATURES Location/Qualifiers
source 1. .29
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Col-0"
/db_xref="taxon:3702"
/clone="SALK_100255.25.15.x"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html"
ORIGIN

Query Match 47.8%; Score 11; DB 8; Length 29;
Best Local Similarity 100.0%; Pred. No. 3.1e+04;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 GCAATTCCGGT 19
|||||||||||
Db 12 GCAATTCCGGT 22

RESULT 3
AZ817376
LOCUS AZ817376 33 bp DNA linear GSS 20-FEB-2001
DEFINITION 2M0086N22R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0086N22 R, genomic survey sequence.
ACCESSION AZ817376
VERSION AZ817376.1 GI:12987380
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 33)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0086 row: N column: 22
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 33.
FEATURES Location/Qualifiers
source 1. .33
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0086N22"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
ORIGIN

Query Match 47.8%; Score 11; DB 8; Length 33;
Best Local Similarity 100.0%; Pred. No. 3.1e+04;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 TGACTGCAATT 14
|||||||||||
Db 23 TGACTGCAATT 33

RESULT 4
BH791854/c
LOCUS BH791854 41 bp DNA linear GSS 02-APR-2002
DEFINITION SALK_061833.40.05.x Arabidopsis thaliana TDNA insertion lines

```
            Arabidopsis thaliana genomic clone SALK_061833.40.05.x, genomic
            survey sequence.
ACCESSION   BH791854
VERSION     BH791854.1  GI:19886147
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 41)
  AUTHORS   Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
            Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
            Shinn,P., Zimmerman,J. and Ecker,J.R.
  TITLE     A Sequence-Indexed Library of Insertion Mutations in the
            Arabidopsis Genome
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Joseph R. Ecker
            Salk Institute Genomic Analysis Laboratory (SIGnAL)
            The Salk Institute for Biological Studies
            10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
            Tel: 858 453 4100 x1752
            Fax: 858 558 6379
            Email: ecker@salk.edu
            This is single pass sequence recovered from the left border of
            TDNA.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .41
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /ecotype="Col-0"
                     /db_xref="taxon:3702"
                     /clone="SALK_061833.40.05.x"
                     /clone_lib="Arabidopsis thaliana TDNA insertion lines"
                     /note="PCR was performed on Arabidopsis thaliana lines
                     each of which contains  one or more TDNA insertion
                     elements. The resultant fragment for each  line was
                     directly sequenced to determine the genomic sequence at
                     the  site of insertion. Details of the protocols used can
                     be found at  http://signal.salk.edu/tdna_protocols.html"
ORIGIN

  Query Match             47.8%;  Score 11;  DB 8;  Length 41;
  Best Local Similarity   100.0%;  Pred. No. 3.1e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 ATGACTGCAAT 13
              ||||||||||
Db         38 ATGACTGCAAT 28


RESULT 5
BH812487/c
LOCUS       BH812487                  41 bp    DNA     linear   GSS 02-MAY-2002
DEFINITION  SALK_061833 Arabidopsis thaliana TDNA insertion lines Arabidopsis
            thaliana genomic clone SALK_061833, genomic survey sequence.
ACCESSION   BH812487
VERSION     BH812487.1  GI:20390942
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 41)
  AUTHORS   Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
            Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
            Shinn,P., Zimmerman,J. and Ecker,J.R.
  TITLE     A Sequence-Indexed Library of Insertion Mutations in the
            Arabidopsis Genome
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Joseph R. Ecker
            Salk Institute Genomic Analysis Laboratory (SIGnAL)
            The Salk Institute for Biological Studies
            10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
            Tel: 858 453 4100 x1752
            Fax: 858 558 6379
            Email: ecker@salk.edu
            This is single pass sequence recovered from the left border of
            TDNA.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .41
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /ecotype="Col-0"
                     /db_xref="taxon:3702"
                     /clone="SALK_061833"
                     /clone_lib="Arabidopsis thaliana TDNA insertion lines"
                     /note="PCR was performed on Arabidopsis thaliana lines
                     each of which contains  one or more TDNA insertion
                     elements. The resultant fragment for each  line was
                     directly sequenced to determine the genomic sequence at
                     the  site of insertion. Details of the protocols used can
                     be found at  http://signal.salk.edu/tdna_protocols.html"
ORIGIN

  Query Match             47.8%;  Score 11;  DB 8;  Length 41;
  Best Local Similarity   100.0%;  Pred. No. 3.1e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 ATGACTGCAAT 13
              ||||||||||
Db         38 ATGACTGCAAT 28


RESULT 6
BG153713/c
LOCUS       BG153713                  50 bp    mRNA    linear   EST 05-FEB-2001
DEFINITION  nag57g06.x1 NCI_CGAP_Co26 Homo sapiens cDNA clone IMAGE:4225738 3',
            mRNA sequence.
ACCESSION   BG153713
VERSION     BG153713.1  GI:12665743
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 50)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .50
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4225738"
                     /tissue_type="normal colonic mucosa"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Co26"
                     /note="Organ: colon; Vector: pAMP1; mRNA made from normal
                     colonic mucosa, cDNA made by oligo-dT priming.
                     Directionally cloned into UDG sites. Size-selected on
                     agarose gel,  average insert size 300 bp. Primary library.
```

cDNA Library Preparation: David B. Krizman, Ph.D.
Reference: Krizman et al. (1996) Cancer Research
56:5380-5383."
ORIGIN

Query Match 47.8%; Score 11; DB 4; Length 50;
Best Local Similarity 100.0%; Pred. No. 3.2e+04;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CATGACTGCAA 12
||||||||||
Db 13 CATGACTGCAA 3

RESULT 7
AW249907/c
LOCUS AW249907 52 bp mRNA linear EST 07-JAN-2000
DEFINITION 2821659.3prime NIH_MGC_7 Homo sapiens cDNA clone IMAGE:2821659 3', mRNA sequence.
ACCESSION AW249907
VERSION AW249907.1 GI:6592900
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 52)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Other_ESTs: 2821659.5prime
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP cDNA Library Preparation: Ling Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Berkeley MGC sequencing project Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html Base Calling / Quality Scores: PHRED from University of Washington Genome Center. Vector Trimming: cross_match from University of Washington Genome Center PHRAP suite. Poly-T Identification: patMatch.pl from Berkeley Drosophila Genome Project. University of Washington Genome Center: http://www.genome.washington.edu Low Quality Sequence: 30 contiguous PHRED high quality bases following vector sequence. Very Low Quality Sequence: Trace file contained 52 contiguous distinct peaks following vector sequence. Polyadenylation: Based upon the presence of a XhoI site followed by a run of 14 or more T residues at the beginning of the sequence, this cDNA insert was polyadenylated.
Plate: LLCM7 row: G column: 4
High quality sequence stop: 30.
FEATURES Location/Qualifiers
source 1. .52
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2821659"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_7"
/note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
ORIGIN

Query Match 47.8%; Score 11; DB 2; Length 52;
Best Local Similarity 100.0%; Pred. No. 3.2e+04;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 ACTGCAATTCC 16
||||||||||
Db 41 ACTGCAATTCC 31

RESULT 8
CN564600/c
LOCUS CN564600 55 bp mRNA linear EST 03-MAY-2004
DEFINITION tag20b04.x1 Hydra EST -Kiel 1 Hydra magnipapillata cDNA 3', mRNA sequence.
ACCESSION CN564600
VERSION CN564600.1 GI:46973904
KEYWORDS EST.
SOURCE Hydra magnipapillata
ORGANISM Hydra magnipapillata
Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Anthomedusae; Hydridae; Hydra.
REFERENCE 1 (bases 1 to 55)
AUTHORS Bode,H., Blumberg,B., Steele,R., Wigge,P., Gee,L., Nguyen,Q., Martinez,D., Kibler,D., Hampson,S., Clifton,S., Pape,D., Marra,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Ronko,I., Tsagareishvili,R., Maguire,L., Kennedy,S., Waterston,R. and Wilson,R.
TITLE WashU Hydra EST Project
JOURNAL Unpublished (2002)
COMMENT Contact: Hans Bode
WashU Hydra EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library was constructed by Konstantin Khalturin, Zoologisches Institut, Univ. Kiel, Germany Library materials provided by Thomas Bosch, Zoologisches Institut, CAU Kiel, Germany DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Hans Bode (hrbode@uci.edu)
Seq primer: degenerate primer.
FEATURES Location/Qualifiers
source 1. .55
/organism="Hydra magnipapillata"
/mol_type="mRNA"
/strain="105"
/db_xref="taxon:6085"
/lab_host="DH5a"
/clone_lib="Hydra EST -Kiel 1"
/note="Vector: pSPORT1; Site_1: Not I; Site_2: Sal I; pSPORT 1 Vector is ampicillin resistent, M13 reverse primer was used by us for sequencing of 5' parts of inserts; 3' parts of cDNAs contain long polyA tracks which makes sequencing from 3' direction complicated"
ORIGIN

Query Match 47.8%; Score 11; DB 7; Length 55;
Best Local Similarity 100.0%; Pred. No. 3.2e+04;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 TTCCGGTCTTT 23
||||||||||
Db 37 TTCCGGTCTTT 27

RESULT 9
BH918919/c
LOCUS BH918919 57 bp DNA linear GSS 12-SEP-2002
DEFINITION 3526_1_63_1_A10.2EL_x_1 3526 - RescueMu Grid K Zea mays genomic, genomic survey sequence.
ACCESSION BH918919
VERSION BH918919.1 GI:22808353
KEYWORDS GSS.

```
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 57)
  AUTHORS   Walbot,V.
  TITLE     Maize genomic sequences found using engineered RescueMu transposon
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Possible ligation site of ends cut by 2 different endonucleases.
            Reverse complemented post-ligation sequence from source sequence.
            Plate: 3526_1_63_1  row: 9
            Class: transposon-tagged.
FEATURES             Location/Qualifiers
     source          1. .57
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="mixed background W23/A188/B73"
                     /db_xref="taxon:4577"
                     /tissue_type="leaf"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="3526 - RescueMu Grid K"
                     /note="Organ: leaf; Vector: RescueMu (engineered from
                     pBlueScript backbone); Site_1: BamHI; Site_2: BglII;
                     RescueMu is a 4.9 kb, modified maize Mu transposon
                     designed to allow plasmid rescue from total genomic DNA.
                     Mu elements insert preferentially into transcription
                     units. For more information on RescueMu, go to the web
                     site 'www.zmdb.iastate.edu' and follow the links for
                     'RescueMu.' Grid K was grown at Molokai, Hawaii in Winter
                     2000-2001. DNA was extracted from leaf punches, double
                     digested using BamHI and BglII, and ligated to form
                     circular plasmids. DH10B cells were transformed and then
                     screened on LB plates with ampicillin."
ORIGIN

  Query Match             47.8%;  Score 11;  DB 8;  Length 57;
  Best Local Similarity   100.0%;  Pred. No. 3.2e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CATGACTGCAA 12
              ||||||||||
Db         51 CATGACTGCAA 41


RESULT 10
AJ588850
LOCUS       AJ588850                19 bp    DNA     linear   GSS 15-JAN-2004
DEFINITION  Arabidopsis thaliana T-DNA flanking sequence, left border, clone
            358F07, genomic survey sequence.
ACCESSION   AJ588850
VERSION     AJ588850.1  GI:37938474
KEYWORDS    GSS; left border; T-DNA flanking sequence.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Brunaud,V., Balzergue,S., Dubreucq,B., Aubourg,S., Samson,F.,
            Chauvin,S., Bechtold,N., Cruaud,C., DeRose,R., Pelletier,G.,
            Lepiniec,L., Caboche,M. and Lecharny,A.
  TITLE     T-DNA integration into the Arabidopsis genome depends on sequences
            of pre-insertion sites
  JOURNAL   EMBO Rep. 3 (12), 1152-1157 (2002)
  MEDLINE   22363535
   PUBMED   12446565
REFERENCE   2  (bases 1 to 19)
  AUTHORS   Balzergue,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2003) Balzergue S., UMRGV, INRA/CNRS, 2 rue
            Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT     PCR was performed on DNA from transformants of Arabidopsis thaliana
            plants from INRA (Versailles). The DNA fragment(s) resulting from
            the PCR were directly sequenced from the left or the right border
            to determine the genomic sequence flanking the insertion. T-DNA
            derived sequences were removed. Information to order the
            corresponding mutant line and a link to a database providing a
            graphical display of the insertion site are available at
            http://dbsgap.versailles.inra.fr/publiclines/. This sequence has
            been generated in the framework of the French plant genomics
            program 'Genoplante' (http://www.genoplante.com and
            http://genoplante-info.infobiogen.fr).
FEATURES             Location/Qualifiers
     source          1. .19
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /cultivar="Wassillewskija"
                     /db_xref="taxon:3702"
                     /clone="358F07"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
     misc_feature    1. .19
                     /note="T-DNA flanking sequence
                     left border"
ORIGIN

  Query Match             43.5%;  Score 10;  DB 9;  Length 19;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+05;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 TGACTGCAAT 13
              ||||||||||
Db          3 TGACTGCAAT 12


RESULT 11
AG197078
LOCUS       AG197078                25 bp    DNA     linear   GSS 06-MAR-2004
DEFINITION  Pan troglodytes DNA, clone: RP43-077A10.TJ, genomic survey
            sequence.
ACCESSION   AG197078
VERSION     AG197078.1  GI:45229254
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Park,H., Kim,Y., Kim,S., Han,Y., Woo,T., Park,K., Eun,C.J.,
            Hoon,S.T., Chu,M., Kim,H., Joo,S., Kim,C., Song,W. and Yoo,H.
  TITLE     BAC end sequences of Library RP-43
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 25)
  AUTHORS   Park,H., Kim,Y., Kim,S., Han,Y., Woo,T., Park,K., Eun,C.J.,
            Hoon,S.T., Chu,M., Kim,H., Joo,S., Kim,C., Song,W. and Yoo,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JAN-2002) Hong-Seog Park, Korea Research Institute of
            Bioscience and Biotechnology (KRIBB), Genome Research Center (GRC);
            52, Oun-dong, Yusong-gu, Daejeon 305-333, Korea
            (E-mail:redstone@mail.kribb.re.kr, URL:http://phs.grc.kribb.re.kr/,
            Tel:82-42-866-7181, Fax:82-42-860-4409)
COMMENT     Clones are derived from the chimpanzee BAC library RP-43 This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
             PRIMERS
                 Sequencing: TJ
             LIBRARY
                 Vector    : pBACe3.6
```

```
                R.Site 1  : EcoRI
                R.Site 2  : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .25
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-077A10.TJ"
                     /sex="male"
                     /cell_type="lymphocytes"
                     /clone_lib="RP-43 Chimpanzee Male BAC Library"
ORIGIN

  Query Match             43.5%;  Score 10;  DB 9;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+05;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          7 CTGCAATTCC 16
              ||||||||||
Db          1 CTGCAATTCC 10


RESULT 12
BX660142
LOCUS       BX660142                 34 bp    DNA     linear   GSS 04-APR-2004
DEFINITION  Arabidopsis thaliana T-DNA flanking sequence GK-650H01-021296,
            genomic survey sequence.
ACCESSION   BX660142
VERSION     BX660142.1  GI:37616530
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Li,Y., Rosso,M.G., Strizhov,N., Viehoever,P. and Weisshaar,B.
  TITLE     GABI-Kat SimpleSearch: a flanking sequence tag (FST) database for
            the identification of T-DNA insertion mutants in Arabidopsis
            thaliana
  JOURNAL   Bioinformatics 19 (11), 1441-1442 (2003)
  MEDLINE   22755829
   PUBMED   12874060
REFERENCE   2
  AUTHORS   Rosso,M.G., Li,Y., Strizhov,N., Reiss,B., Dekker,K. and
            Weisshaar,B.
  TITLE     An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for
            flanking sequence tag-based reverse genetics
  JOURNAL   Plant Mol. Biol. 53 (1-2), 247-259 (2003)
  MEDLINE   23117147
   PUBMED   14756321
REFERENCE   3
  AUTHORS   Strizhov,N., Li,Y., Rosso,M.G., Viehoever,P., Dekker,K.A. and
            Weisshaar,B.
  TITLE     High-throughput generation of sequence indexes from T-DNA
            mutagenized Arabidopsis thaliana lines
  JOURNAL   BioTechniques 35 (6), 1164-1168 (2003)
   PUBMED   14682050
REFERENCE   4  (bases 1 to 34)
  AUTHORS   Li,Y., Strizhov,N., Rosso,M.G. and Weisshaar,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2004) Weisshaar B., Max-Planck-Institut fuer
            Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
COMMENT     This sequence has been recovered from the left border of the T-DNA.
            It indicates an insertion within the locus defined by BAC clone
            T19E23. Details on the protocols used for generation of the
            sequence are described in References 1-3. The sequences are
            generated at the MPI for Plant Breeding Research in the context of
            the GABI-Kat project. GABI-Kat is part of the German Plant Genomics
            program designated 'GABI'. Information on line availability can be
            found at: http://www.mpiz-koeln.mpg.de/GABI-Kat/.
FEATURES             Location/Qualifiers
     source          1. .34
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia 0"
                     /db_xref="taxon:3702"
                     /clone="GK-650H01-021296"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
                     /ecotype="Col-0"
                     /note="PCR was performed on DNA from Arabidopsis thaliana
                     plants (T1) which were transformed with the T-DNA from
                     vector pAC161 (GenBank accession number: AJ537514). The
                     lines contain one or more T-DNA insertions. The DNA
                     fragment(s) resulting from the PCR were directly sequenced
                     to determine the genomic sequence flanking the insertion.
                     T-DNA derived sequences were removed."
ORIGIN

  Query Match             43.5%;  Score 10;  DB 9;  Length 34;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+05;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 TGACTGCAAT 13
              ||||||||||
Db         17 TGACTGCAAT 26


RESULT 13
BH810737/c
LOCUS       BH810737                 36 bp    DNA     linear   GSS 02-MAY-2002
DEFINITION  SALK_051126 Arabidopsis thaliana TDNA insertion lines Arabidopsis
            thaliana genomic clone SALK_051126, genomic survey sequence.
ACCESSION   BH810737
VERSION     BH810737.1  GI:20388555
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 36)
  AUTHORS   Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
            Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
            Shinn,P., Zimmerman,J. and Ecker,J.R.
  TITLE     A Sequence-Indexed Library of Insertion Mutations in the
            Arabidopsis Genome
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Joseph R. Ecker
            Salk Institute Genomic Analysis Laboratory (SIGnAL)
            The Salk Institute for Biological Studies
            10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
            Tel: 858 453 4100 x1752
            Fax: 858 558 6379
            Email: ecker@salk.edu
            This is single pass sequence recovered from the left border of
            TDNA.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .36
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /ecotype="Col-0"
                     /db_xref="taxon:3702"
                     /clone="SALK_051126"
                     /clone_lib="Arabidopsis thaliana TDNA insertion lines"
                     /note="PCR was performed on Arabidopsis thaliana lines
                     each of which contains  one or more TDNA insertion
                     elements. The resultant fragment for each  line was
                     directly sequenced to determine the genomic sequence at
                     the  site of insertion. Details of the protocols used can
                     be found at  http://signal.salk.edu/tdna_protocols.html"
ORIGIN

  Query Match             43.5%;  Score 10;  DB 8;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+05;
```

```
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          14 TCCGGTCTTT 23
               ||||||||||
Db          18 TCCGGTCTTT 9


RESULT 14
DME546528/c
LOCUS       DME546528                36 bp    DNA     linear   GSS 24-FEB-2003
DEFINITION  Drosophila melanogaster flanking sequence of RS P element insertion
            P{RS5}5-HA-1904, clone library P{RS5}, genomic survey sequence.
ACCESSION   AJ546528
VERSION     AJ546528.1  GI:28554603
KEYWORDS    GSS; genome survey sequence.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1
  AUTHORS   Ryder,E.J., Ashburner,M., Bagunya,J., Blows,F., Bucheton,A.,
            Coulson,D., Dickson,B., Drummond,J., Glover,D., Gunton,N.,
            Hafen,E., Hall,S., Heisenberg,M., Lepesant,J.A., Maroy,P.,
            Mechler,B., O'Kane,C., Pflugfelder,G., Rasmuson-Lestander,A.,
            Reuter,G., Roote,J., Szidonya,J., Wang,S., Webster,J. and
            Russell,S.
  TITLE     Mapping of RS P element insertions in Drosophila melanogaster for
            the DrosDel second generation deficiency kit
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 36)
  AUTHORS   Ryder,E.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-2003) Ryder E.J., Department of Genetics,
            University of Cambridge, Downing Street, CB2 3EH, UNITED KINGDOM
COMMENT     The insertion point of the P element is before base 1 of the
            sequence. Further information about this P element insertion line
            can be found at http://www.flyseq.org.uk and
            http://www.drosdel.org.uk.
FEATURES             Location/Qualifiers
     source          1. .36
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /chromosome="2L"
                     /clone="P{RS5}5-HA-1904"
                     /clone_lib="P{RS5}"
                     /note="read=5' end"
     misc_feature    1. .36
                     /note="P element insertion in the 3' to 5' orientation"
ORIGIN

  Query Match             43.5%;  Score 10;  DB 9;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+05;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          6 ACTGCAATTC 15
              ||||||||||
Db         23 ACTGCAATTC 14


RESULT 15
AA972482
LOCUS       AA972482                 37 bp    mRNA    linear   EST 07-JUL-1998
DEFINITION  op42d03.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:1579493 3' similar to TR:Q13526 Q13526 PIN1. ;, mRNA
            sequence.
ACCESSION   AA972482
VERSION     AA972482.1  GI:3145246
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 37)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Trace considered overall poor quality
            Insert Length: 523   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .37
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1579493"
                     /lab_host="DH10B"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo.  "
ORIGIN

  Query Match             43.5%;  Score 10;  DB 1;  Length 37;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+05;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CATGACTGCA 11
              ||||||||||
Db          1 CATGACTGCA 10


Search completed: September  6, 2005, 23:10:13
Job time : 1639 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 6, 2005, 20:30:00 ; Search time 233 Seconds
(without alignments)
584.352 Million cell updates/sec

Title: US-10-729-421-8
Perfect score: 23
Sequence: 1 tcatgactgcaattccggtcttt 23

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 4390206 seqs, 2959870667 residues

Word size : 10

Total number of hits satisfying chosen parameters: 684

Minimum DB seq length: 0
Maximum DB seq length: 60

Post-processing: Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	23	100.0	23	12	ADQ30638	Adq30638 West Nile
	2	23	100.0	46	12	ADQ30655	Adq30655 West Nile
c	3	20	87.0	36	8	ABZ25451	Abz25451 PCR prime
c	4	20	87.0	36	9	AAL55873	Aal55873 FC117 PCR
c	5	20	87.0	36	12	ADM16871	Adm16871 Plasmid p
c	6	20	87.0	39	8	ABZ25450	Abz25450 West Nile
c	7	20	87.0	39	9	AAL55872	Aal55872 FC116 RT-
c	8	20	87.0	39	12	ADM16870	Adm16870 West Nile
c	9	20	87.0	51	10	ADC06634	Adc06634 PCR prime
c	10	20	87.0	51	10	ADC06633	Adc06633 PCR prime
c	11	20	87.0	58	12	ADM16878	Adm16878 Plasmid p
c	12	19	82.6	33	8	ABZ25440	Abz25440 PCR prime
c	13	19	82.6	33	9	AAL55862	Aal55862 FC110 PCR
c	14	19	82.6	33	12	ADM16860	Adm16860 Plasmid p
c	15	19	82.6	36	8	ABZ25431	Abz25431 West Nile
c	16	19	82.6	36	9	AAL57662	Aal57662 CF107 RT-
c	17	19	82.6	36	12	ADM16851	Adm16851 West Nile
c	18	19	82.6	51	10	ADC06635	Adc06635 PCR prime
c	19	19	82.6	56	4	AAD14623	Aad14623 WN virus-
	20	17	73.9	17	6	ACN09412	Acn09412 WNV minus
c	21	17	73.9	17	6	ACN00067	Acn00067 WNV Hamme
	22	17	73.9	17	6	ACN13570	Acn13570 WNV minus
c	23	17	73.9	17	6	ACN03509	Acn03509 WNV Zinzy
	24	17	73.9	17	6	ACN14150	Acn14150 WNV minus
c	25	17	73.9	17	6	ACN05528	Acn05528 WNV Amber
	26	17	73.9	17	6	ACN12169	Acn12169 WNV minus
c	27	17	73.9	17	6	ACN04748	Acn04748 WNV DNAzy
	28	17	73.9	17	6	ACN09413	Acn09413 WNV minus
	29	16	69.6	17	6	ACN12168	Acn12168 WNV minus
c	30	16	69.6	17	6	ACN05527	Acn05527 WNV Amber
	31	16	69.6	17	6	ACN12170	Acn12170 WNV minus
c	32	16	69.6	17	6	ACN01482	Acn01482 WNV Inozy
c	33	16	69.6	30	10	ADC06605	Adc06605 WNV 1st j
c	34	15	65.2	17	6	ACN03510	Acn03510 WNV Zinzy
	35	15	65.2	17	6	ACN15245	Acn15245 WNV minus
c	36	15	65.2	17	6	ACN01481	Acn01481 WNV Inozy
	37	14	60.9	17	6	ACN14149	Acn14149 WNV minus
	38	14	60.9	17	6	ACN13571	Acn13571 WNV minus
c	39	13	56.5	17	6	ACN00068	Acn00068 WNV Hamme
c	40	13	56.5	17	6	ACN04747	Acn04747 WNV DNAzy
	41	13	56.5	17	6	ACN15244	Acn15244 WNV minus
c	42	13	56.5	25	9	ACK26449	Ack26449 Human mic
c	43	13	56.5	30	10	ADC06609	Adc06609 Chimeric
c	44	13	56.5	30	10	ADC06611	Adc06611 Chimeric
	45	12	52.2	17	6	ACN09414	Acn09414 WNV minus

ALIGNMENTS

RESULT 1
ADQ30638
ID ADQ30638 standard; DNA; 23 BP.
XX
AC ADQ30638;
XX
DT 23-SEP-2004 (first entry)
XX
DE West Nile Virus capture oligonucleotide WNVVC8.
XX
KW ss; capture oligonucleotide; West Nile Virus; diagnosis.
XX
OS West Nile virus.
XX
PN WO2004055159-A2.
XX
PD 01-JUL-2004.
XX
PF 05-DEC-2003; 2003WO-US038750.
XX
PR 12-DEC-2002; 2002US-0432850P.
PR 20-JUN-2003; 2003US-0480431P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Shyamala V;
XX
DR WPI; 2004-488058/46.
XX
PT New isolated oligonucleotides for accurately diagnosing West Nile virus
PT infection or for capturing, detecting and quantitating West Nile virus in
PT blood samples.
XX
PS Claim 1; SEQ ID NO 8; 56pp; English.
XX
CC The invention relates to an isolated oligonucleotide not more than 60
CC nucleotides in length comprising a nucleotide sequence (S1) of at least
CC 10 contiguous nucleotides from any of the 28 nucleotide sequences (e.g.
CC 20, 21 or 23 bp) given in the specification derived from the West Nile
CC Virus (WNV) genome, a nucleotide sequence (S2) having 90% sequence
CC identity to the nucleotide sequence of (S1), or complements of (S1) and
CC (S2). The oligonucleotide further comprises a detectable label at the 5'-
CC end and/or the 3'-end. The detectable label is a fluorescent label

```
CC   selected from 6-carboxyfluorescein (6-FAM), tetramethyl rhodamine
CC   (TAMRA), and 2',4',5',7',-tetrachloro-4-7-dichlorofluorescein (TET). The
CC   composition and methods are useful for accurately diagnosing West Nile
CC   virus infection or for capturing, detecting and quantitating West Nile
CC   virus in biological samples, particularly blood samples. This sequence
CC   corresponds to a capture oligonucleotide of the invention.
XX
SQ   Sequence 23 BP; 4 A; 6 C; 4 G; 9 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 23;  DB 12;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 0.00091;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATGACTGCAATTCCGGTCTTT 23
              |||||||||||||||||||||||
Db          1 TCATGACTGCAATTCCGGTCTTT 23


RESULT 2
ADQ30655
ID   ADQ30655 standard; DNA; 46 BP.
XX
AC   ADQ30655;
XX
DT   23-SEP-2004  (first entry)
XX
DE   West Nile Virus capture oligonucleotide poly-A-WNVVC8.
XX
KW   ss; capture oligonucleotide; West Nile Virus; diagnosis.
XX
OS   West Nile virus.
XX
PN   WO2004055159-A2.
XX
PD   01-JUL-2004.
XX
PF   05-DEC-2003; 2003WO-US038750.
XX
PR   12-DEC-2002; 2002US-0432850P.
PR   20-JUN-2003; 2003US-0480431P.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Shyamala V;
XX
DR   WPI; 2004-488058/46.
XX
PT   New isolated oligonucleotides for accurately diagnosing West Nile virus
PT   infection or for capturing, detecting and quantitating West Nile virus in
PT   blood samples.
XX
PS   Example 1; SEQ ID NO 25; 56pp; English.
XX
CC   The invention relates to an isolated oligonucleotide not more than 60
CC   nucleotides in length comprising a nucleotide sequence (S1) of at least
CC   10 contiguous nucleotides from any of the 28 nucleotide sequences (e.g.
CC   20, 21 or 23 bp) given in the specification derived from the West Nile
CC   Virus (WNV) genome, a nucleotide sequence (S2) having 90% sequence
CC   identity to the nucleotide sequence of (S1), or complements of (S1) and
CC   (S2). The oligonucleotide further comprises a detectable label at the 5'-
CC   end and/or the 3'-end. The detectable label is a fluorescent label
CC   selected from 6-carboxyfluorescein (6-FAM), tetramethyl rhodamine
CC   (TAMRA), and 2',4',5',7',-tetrachloro-4-7-dichlorofluorescein (TET). The
CC   composition and methods are useful for accurately diagnosing West Nile
CC   virus infection or for capturing, detecting and quantitating West Nile
CC   virus in biological samples, particularly blood samples. This sequence
CC   corresponds to a capture oligonucleotide of the invention.
XX
SQ   Sequence 46 BP; 27 A; 6 C; 4 G; 9 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 23;  DB 12;  Length 46;
  Best Local Similarity   100.0%;  Pred. No. 0.00088;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATGACTGCAATTCCGGTCTTT 23
              |||||||||||||||||||||||
Db         24 TCATGACTGCAATTCCGGTCTTT 46


RESULT 3
ABZ25451/c
ID   ABZ25451 standard; DNA; 36 BP.
XX
AC   ABZ25451;
XX
DT   27-MAR-2003  (first entry)
XX
DE   PCR primer FC117, SEQ ID 29.
XX
KW   Virucide; vaccine; horse; dog; cat; cattle; pig; bird; West Nile virus;
KW   WNV; PCR; primer; ss.
XX
OS   Synthetic.
XX
PN   WO200281621-A2.
XX
PD   17-OCT-2002.
XX
PF   05-APR-2002; 2002WO-FR001200.
XX
PR   06-APR-2001; 2001FR-00004737.
XX
PA   (MERI-) MERIAL.
XX
PI   Loosmore SM,  Audonnet JF;
XX
DR   WPI; 2003-111799/10.
XX
PT   Vaccine for treatment or prevention of West Nile virus (WNV) infection,
PT   for use in veterinary medicine, comprises a recombinant virus expresssing
PT   a WNV structural protein.
XX
PS   Example 18; Page 41; 56pp; French.
XX
CC   The present invention relates to a vaccine for protecting horses, dogs,
CC   cats, cattle, pigs and birds against West Nile virus (WNV). The vaccine
CC   comprises: (i) one or more recombinant avipox, NYVAC or MVA viruses that
CC   express one of the WNV proteins prM, M and E and (ii) a vehicle or
CC   excipient. The present sequence is a PCR primer, which was used in an
CC   example from the invention
XX
SQ   Sequence 36 BP; 8 A; 7 C; 10 G; 11 T; 0 U; 0 Other;

  Query Match             87.0%;  Score 20;  DB 8;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 0.052;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATGACTGCAATTCCGGTC 20
              ||||||||||||||||||||
Db         31 TCATGACTGCAATTCCGGTC 12


RESULT 4
AAL55873/c
ID   AAL55873 standard; DNA; 36 BP.
XX
AC   AAL55873;
XX
DT   06-NOV-2003  (first entry)
XX
DE   FC117 PCR primer used to amplify the plasmid pFC115.
XX
KW   Immunogenic composition; West nile fever virus; WNV; prM; M; membrane; E;
KW   pre-membrane protein; envelope; virucide; vaccine; FC117; primer; PCR;
KW   ss; plasmid pFC115.
XX
```

OS Unidentified.
XX
PN US2003104008-A1.
XX
PD 05-JUN-2003.
XX
PF 04-APR-2002; 2002US-00116298.
XX
PR 06-APR-2001; 2001US-0281923P.
XX
PA (LOOS/) LOOSMORE S M.
PA (AUDO/) AUDONNET J F.
XX
PI Loosmore SM, Audonnet JF;
XX
DR WPI; 2003-567944/53.
XX
PT New immunogenic composition comprising a recombinant avipox virus that
PT expresses in vivo in the animal the West Nile (WN) proteins prM, M or E,
PT useful for inducing an immunological response against WN virus.
XX
PS Example 18; Page 14; 24pp; English.
XX
CC The invention relates to a novel immunogenic composition for inducing an
CC immune response against West Nile fever virus (WNV) in an animal. The
CC composition comprises a vehicle or excipient and a recombinant avipox
CC virus that expresses in vivo in the animal the WNV proteins prM (pre-
CC membrane protein), M (membrane protein) or E (envelope protein). The
CC animal is selected from canine, feline, bovine, porcine, chicken, equine,
CC a duck, a goose or a turkey. The composition of the invention
CC demonstrates virucide activity and may be useful as a vaccine against
CC WNV. The current sequence is that of the FC117 PCR primer of the
CC invention which was used to amplify the plasmid pFC115
XX
SQ Sequence 36 BP; 8 A; 7 C; 10 G; 11 T; 0 U; 0 Other;

Query Match 87.0%; Score 20; DB 9; Length 36;
Best Local Similarity 100.0%; Pred. No. 0.052;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCATGACTGCAATTCCGGTC 20
||||||||||||||||||||
Db 31 TCATGACTGCAATTCCGGTC 12

RESULT 5
ADM16871/c
ID ADM16871 standard; DNA; 36 BP.
XX
AC ADM16871;
XX
DT 20-MAY-2004 (first entry)
XX
DE Plasmid pFC115 PCR primer #1.
XX
KW immunogen; vaccine; West Nile virus; ss; PCR; primer.
XX
OS Synthetic.
XX
PN US2004037848-A1.
XX
PD 26-FEB-2004.
XX
PF 26-FEB-2003; 2003US-00374953.
XX
PR 06-APR-2001; 2001US-0281923P.
PR 04-APR-2002; 2002US-00116298.
XX
PA (AUDO/) AUDONNET J F.
PA (MINK/) MINKE J M.
PA (LOOS/) LOOSMORE S M.
PA (KARA/) KARACA K.
XX
PI Audonnet JF, Minke JM, Loosmore SM, Karaca K;
XX
DR WPI; 2004-191012/18.
XX
PT Vaccine composition, useful in inducing an immune response against West
PT Nile virus, comprises a vector that contains heterologous nucleic acid
PT molecule(s), and that expresses in vivo in the animal a WNV protein.
XX
PS Example 18; SEQ ID NO 29; 36pp; English.
XX
CC The invention relates to an immunogenic or vaccine composition which
CC induces an immune response against West Nile virus (WNV) in an animal
CC susceptible to WNV comprises a vector that contains heterologous nucleic
CC acid molecule(s) and that expresses in vivo in the animal a WNV E; WNV
CC prM and E; WNV M and E; WNV prM, WNV M and E, WNV polyprotein prM-E, WNV
CC polyprotein M-E, or WNV polyprotein prM-M-E. The composition is useful
CC for inducing an immunological or protective immune response against WNV
CC and against another pathogen of the animal. Also inducing an
CC immunological or protective immune response against WNV in an animal
CC comprises administering to the animal (a) the immunogenic or vaccine
CC composition and (b) a WNV isolated antigen, immunogen or epitope, where
CC (a) is administered prior to (b) in a prime-boost regimen, or (b) is
CC administered prior to (a) in a prime-boost regimen, or (a) and (b) are
CC administered together, either sequentially or in admixture. The present
CC sequence is used in the exemplification of the invention.
XX
SQ Sequence 36 BP; 8 A; 7 C; 10 G; 11 T; 0 U; 0 Other;

Query Match 87.0%; Score 20; DB 12; Length 36;
Best Local Similarity 100.0%; Pred. No. 0.052;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCATGACTGCAATTCCGGTC 20
||||||||||||||||||||
Db 31 TCATGACTGCAATTCCGGTC 12

RESULT 6
ABZ25450/c
ID ABZ25450 standard; DNA; 39 BP.
XX
AC ABZ25450;
XX
DT 27-MAR-2003 (first entry)
XX
DE West Nile Virus PCR primer FC116, SEQ ID 28.
XX
KW Virucide; vaccine; horse; dog; cat; cattle; pig; bird; West Nile virus;
KW WNV; PCR; primer; ss.
XX
OS West Nile Virus.
XX
PN WO200281621-A2.
XX
PD 17-OCT-2002.
XX
PF 05-APR-2002; 2002WO-FR001200.
XX
PR 06-APR-2001; 2001FR-00004737.
XX
PA (MERI-) MERIAL.
XX
PI Loosmore SM, Audonnet JF;
XX
DR WPI; 2003-111799/10.
XX
PT Vaccine for treatment or prevention of West Nile virus (WNV) infection,
PT for use in veterinary medicine, comprises a recombinant virus expresssing
PT a WNV structural protein.
XX
PS Example 17; Page 40; 56pp; French.
XX
CC The present invention relates to a vaccine for protecting horses, dogs,

```
CC   cats, cattle, pigs and birds against West Nile virus (WNV). The vaccine
CC   comprises: (i) one or more recombinant avipox, NYVAC or MVA viruses that
CC   express one of the WNV proteins prM, M and E and (ii) a vehicle or
CC   excipient. The present sequence is a PCR primer, which was used in an
CC   example from the invention
XX
SQ   Sequence 39 BP; 9 A; 6 C; 9 G; 15 T; 0 U; 0 Other;

  Query Match             87.0%;  Score 20;  DB 8;  Length 39;
  Best Local Similarity   100.0%;  Pred. No. 0.052;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATGACTGCAATTCCGGTC 20
              ||||||||||||||||||||
Db         34 TCATGACTGCAATTCCGGTC 15


RESULT 7
AAL55872/c
ID   AAL55872 standard; DNA; 39 BP.
XX
AC   AAL55872;
XX
DT   06-NOV-2003  (first entry)
XX
DE   FC116 RT-PCR primer used to amplify West Nile fever virus RNA.
XX
KW   Immunogenic composition; West nile fever virus; WNV; prM; M; membrane; E;
KW   pre-membrane protein; envelope; virucide; vaccine; FC116; RT-PCR; primer;
KW   PCR; ss.
XX
OS   Unidentified.
OS   Synthetic.
XX
PN   US2003104008-A1.
XX
PD   05-JUN-2003.
XX
PF   04-APR-2002; 2002US-00116298.
XX
PR   06-APR-2001; 2001US-0281923P.
XX
PA   (LOOS/) LOOSMORE S M.
PA   (AUDO/) AUDONNET J F.
XX
PI   Loosmore SM,  Audonnet JF;
XX
DR   WPI; 2003-567944/53.
XX
PT   New immunogenic composition comprising a recombinant avipox virus that
PT   expresses in vivo in the animal the West Nile (WN) proteins prM, M or E,
PT   useful for inducing an immunological response against WN virus.
XX
PS   Example 17; Page 14; 24pp; English.
XX
CC   The invention relates to a novel immunogenic composition for inducing an
CC   immune response against West Nile fever virus (WNV) in an animal. The
CC   composition comprises a vehicle or excipient and a recombinant avipox
CC   virus that expresses in vivo in the animal the WNV proteins prM (pre-
CC   membrane protein), M (membrane protein) or E (envelope protein). The
CC   animal is selected from canine, feline, bovine, porcine, chicken, equine,
CC   a duck, a goose or a turkey. The composition of the invention
CC   demonstrates virucide activity and may be useful as a vaccine against
CC   WNV. The current sequence is that of the FC116 RT-PCR primer of the
CC   invention which was used to amplify West Nile fever virus RNA
XX
SQ   Sequence 39 BP; 9 A; 6 C; 9 G; 15 T; 0 U; 0 Other;

  Query Match             87.0%;  Score 20;  DB 9;  Length 39;
  Best Local Similarity   100.0%;  Pred. No. 0.052;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATGACTGCAATTCCGGTC 20
              ||||||||||||||||||||
Db         34 TCATGACTGCAATTCCGGTC 15


RESULT 8
ADM16870/c
ID   ADM16870 standard; DNA; 39 BP.
XX
AC   ADM16870;
XX
DT   20-MAY-2004  (first entry)
XX
DE   West Nile virus RT-PCR primer #10.
XX
KW   immunogen; vaccine; West Nile virus; ss; reverse transcriptase; RT-PCR;
KW   primer.
XX
OS   West Nile virus.
XX
PN   US2004037848-A1.
XX
PD   26-FEB-2004.
XX
PF   26-FEB-2003; 2003US-00374953.
XX
PR   06-APR-2001; 2001US-0281923P.
PR   04-APR-2002; 2002US-00116298.
XX
PA   (AUDO/) AUDONNET J F.
PA   (MINK/) MINKE J M.
PA   (LOOS/) LOOSMORE S M.
PA   (KARA/) KARACA K.
XX
PI   Audonnet JF,  Minke JM,  Loosmore SM,  Karaca K;
XX
DR   WPI; 2004-191012/18.
XX
PT   Vaccine composition, useful in inducing an immune response against West
PT   Nile virus, comprises a vector that contains heterologous nucleic acid
PT   molecule(s), and that expresses in vivo in the animal a WNV protein.
XX
PS   Example 17; SEQ ID NO 28; 36pp; English.
XX
CC   The invention relates to an immunogenic or vaccine composition which
CC   induces an immune response against West Nile virus (WNV) in an animal
CC   susceptible to WNV comprises a vector that contains heterologous nucleic
CC   acid molecule(s) and that expresses in vivo in the animal a WNV E; WNV
CC   prM and E; WNV M and E; WNV prM, WNV M and E, WNV polyprotein prM-E, WNV
CC   polyprotein M-E, or WNV polyprotein prM-M-E. The composition is useful
CC   for inducing an immunological or protective immune response against WNV
CC   and against another pathogen of the animal. Also inducing an
CC   immunological or protective immune response against WNV in an animal
CC   comprises administering to the animal (a) the immunogenic or vaccine
CC   composition and (b) a WNV isolated antigen, immunogen or epitope, where
CC   (a) is administered prior to (b) in a prime-boost regimen, or (b) is
CC   administered prior to (a) in a prime-boost regimen, or (a) and (b) are
CC   administered together, either sequentially or in admixture. The present
CC   sequence is used in the exemplification of the invention.
XX
SQ   Sequence 39 BP; 9 A; 6 C; 9 G; 15 T; 0 U; 0 Other;

  Query Match             87.0%;  Score 20;  DB 12;  Length 39;
  Best Local Similarity   100.0%;  Pred. No. 0.052;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATGACTGCAATTCCGGTC 20
              ||||||||||||||||||||
Db         34 TCATGACTGCAATTCCGGTC 15


RESULT 9
ADC06634/c
ID   ADC06634 standard; DNA; 51 BP.
```

XX
AC ADC06634;
XX
DT 18-DEC-2003 (first entry)
XX
DE PCR primer SEQ ID 35 used during construction of WNV/DEN4 chimeras.
XX
KW West Nile virus; WNV; DEN4; Dengue virus type 4; virucide; vaccine; ss;
KW PCR; primer.
XX
OS Unidentified.
XX
PN WO2003059384-A1.
XX
PD 24-JUL-2003.
XX
PF 09-JAN-2003; 2003WO-US000594.
XX
PR 10-JAN-2002; 2002US-0347281P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Pletnev AG, Putnak JR, Chanock RM, Murphy BR, Whitehead SS;
PI Blaney JE;
XX
DR WPI; 2003-636686/60.
XX
PT Novel nucleic acid chimera comprising nucleic acids encoding structural
PT protein from West Nile virus and non-structural proteins from wild-type
PT strain of dengue virus useful for producing live West Nile virus
PT vaccines.
XX
PS Disclosure; Page 20; 53pp; English.
XX
CC The invention relates to a novel nucleic acid chimera comprising a first
CC nucleotide sequence encoding at least one structural protein from a West
CC Nile virus (WNV) and a second nucleotide sequence encoding non-structural
CC proteins from a wild-type strain of Dengue virus (DEN), such as Dengue
CC virus type 4 (DEN4). The nucleotide of the invention demonstrates
CC virucide activity and may be useful for producing a WNV vaccine. The
CC current sequence is that of the PCR primer of the invention which was
CC used during the construction of the WNV/DEN4 chimeras.
XX
SQ Sequence 51 BP; 20 A; 9 C; 13 G; 9 T; 0 U; 0 Other;

Query Match 87.0%; Score 20; DB 10; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.051;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCATGACTGCAATTCCGGTC 20
||||||||||||||||||||
Db 46 TCATGACTGCAATTCCGGTC 27

RESULT 10
ADC06633/c
ID ADC06633 standard; DNA; 51 BP.
XX
AC ADC06633;
XX
DT 18-DEC-2003 (first entry)
XX
DE PCR primer SEQ ID 34 used during construction of WNV/DEN4 chimeras.
XX
KW West Nile virus; WNV; DEN4; Dengue virus type 4; virucide; vaccine; ss;
KW PCR; primer.
XX
OS Unidentified.
XX
PN WO2003059384-A1.
XX
PD 24-JUL-2003.
XX
PF 09-JAN-2003; 2003WO-US000594.
XX
PR 10-JAN-2002; 2002US-0347281P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Pletnev AG, Putnak JR, Chanock RM, Murphy BR, Whitehead SS;
PI Blaney JE;
XX
DR WPI; 2003-636686/60.
XX
PT Novel nucleic acid chimera comprising nucleic acids encoding structural
PT protein from West Nile virus and non-structural proteins from wild-type
PT strain of dengue virus useful for producing live West Nile virus
PT vaccines.
XX
PS Disclosure; Page 19; 53pp; English.
XX
CC The invention relates to a novel nucleic acid chimera comprising a first
CC nucleotide sequence encoding at least one structural protein from a West
CC Nile virus (WNV) and a second nucleotide sequence encoding non-structural
CC proteins from a wild-type strain of Dengue virus (DEN), such as Dengue
CC virus type 4 (DEN4). The nucleotide of the invention demonstrates
CC virucide activity and may be useful for producing a WNV vaccine. The
CC current sequence is that of the PCR primer of the invention which was
CC used during the construction of the WNV/DEN4 chimeras.
XX
SQ Sequence 51 BP; 20 A; 9 C; 12 G; 10 T; 0 U; 0 Other;

Query Match 87.0%; Score 20; DB 10; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.051;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCATGACTGCAATTCCGGTC 20
||||||||||||||||||||
Db 46 TCATGACTGCAATTCCGGTC 27

RESULT 11
ADM16878/c
ID ADM16878 standard; DNA; 58 BP.
XX
AC ADM16878;
XX
DT 20-MAY-2004 (first entry)
XX
DE Plasmid pFC115 PCR primer #2.
XX
KW immunogen; vaccine; West Nile virus; ss; PCR; primer.
XX
OS Synthetic.
XX
PN US2004037848-A1.
XX
PD 26-FEB-2004.
XX
PF 26-FEB-2003; 2003US-00374953.
XX
PR 06-APR-2001; 2001US-0281923P.
PR 04-APR-2002; 2002US-00116298.
XX
PA (AUDO/) AUDONNET J F.
PA (MINK/) MINKE J M.
PA (LOOS/) LOOSMORE S M.
PA (KARA/) KARACA K.
XX
PI Audonnet JF, Minke JM, Loosmore SM, Karaca K;
XX
DR WPI; 2004-191012/18.
XX
PT Vaccine composition, useful in inducing an immune response against West
PT Nile virus, comprises a vector that contains heterologous nucleic acid
PT molecule(s), and that expresses in vivo in the animal a WNV protein.

```
XX
PS   Example 28; SEQ ID NO 36; 36pp; English.
XX
CC   The invention relates to an immunogenic or vaccine composition which
CC   induces an immune response against West Nile virus (WNV) in an animal
CC   susceptible to WNV comprises a vector that contains heterologous nucleic
CC   acid molecule(s) and that expresses in vivo in the animal a WNV E; WNV
CC   prM and E; WNV M and E; WNV prM, WNV M and E, WNV polyprotein prM-E, WNV
CC   polyprotein M-E, or WNV polyprotein prM-M-E. The composition is useful
CC   for inducing an immunological or protective immune response against WNV
CC   and against another pathogen of the animal. Also inducing an
CC   immunological or protective immune response against WNV in an animal
CC   comprises administering to the animal (a) the immunogenic or vaccine
CC   composition and (b) a WNV isolated antigen, immunogen or epitope, where
CC   (a) is administered prior to (b) in a prime-boost regimen, or (b) is
CC   administered prior to (a) in a prime-boost regimen, or (a) and (b) are
CC   administered together, either sequentially or in admixture. The present
CC   sequence is used in the exemplification of the invention.
XX
SQ   Sequence 58 BP; 13 A; 10 C; 14 G; 21 T; 0 U; 0 Other;

  Query Match             87.0%;  Score 20;  DB 12;  Length 58;
  Best Local Similarity   100.0%;  Pred. No. 0.051;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATGACTGCAATTCCGGTC 20
              ||||||||||||||||||||
Db         53 TCATGACTGCAATTCCGGTC 34


RESULT 12
ABZ25440/c
ID   ABZ25440 standard; DNA; 33 BP.
XX
AC   ABZ25440;
XX
DT   27-MAR-2003  (first entry)
XX
DE   PCR primer FC110, SEQ ID 18.
XX
KW   Virucide; vaccine; horse; dog; cat; cattle; pig; bird; West Nile virus;
KW   WNV; PCR; primer; ss.
XX
OS   Synthetic.
XX
PN   WO200281621-A2.
XX
PD   17-OCT-2002.
XX
PF   05-APR-2002; 2002WO-FR001200.
XX
PR   06-APR-2001; 2001FR-00004737.
XX
PA   (MERI-) MERIAL.
XX
PI   Loosmore SM,  Audonnet JF;
XX
DR   WPI; 2003-111799/10.
XX
PT   Vaccine for treatment or prevention of West Nile virus (WNV) infection,
PT   for use in veterinary medicine, comprises a recombinant virus expresssing
PT   a WNV structural protein.
XX
PS   Example 9; Page 34; 56pp; French.
XX
CC   The present invention relates to a vaccine for protecting horses, dogs,
CC   cats, cattle, pigs and birds against West Nile virus (WNV). The vaccine
CC   comprises: (i) one or more recombinant avipox, NYVAC or MVA viruses that
CC   express one of the WNV proteins prM, M and E and (ii) a vehicle or
CC   excipient. The present sequence is a PCR primer, which was used in an
CC   example from the invention
XX
SQ   Sequence 33 BP; 7 A; 7 C; 9 G; 10 T; 0 U; 0 Other;

  Query Match             82.6%;  Score 19;  DB 8;  Length 33;
  Best Local Similarity   100.0%;  Pred. No. 0.2;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATGACTGCAATTCCGGT 19
              |||||||||||||||||||
Db         28 TCATGACTGCAATTCCGGT 10


RESULT 13
AAL55862/c
ID   AAL55862 standard; DNA; 33 BP.
XX
AC   AAL55862;
XX
DT   06-NOV-2003  (first entry)
XX
DE   FC110 PCR primer used to amplify the plasmid pFC105.
XX
KW   Immunogenic composition; West nile fever virus; WNV; prM; M; membrane; E;
KW   pre-membrane protein; envelope; virucide; vaccine; FC110; primer; PCR;
KW   ss; plasmid pFC105.
XX
OS   Unidentified.
XX
PN   US2003104008-A1.
XX
PD   05-JUN-2003.
XX
PF   04-APR-2002; 2002US-00116298.
XX
PR   06-APR-2001; 2001US-0281923P.
XX
PA   (LOOS/) LOOSMORE S M.
PA   (AUDO/) AUDONNET J F.
XX
PI   Loosmore SM,  Audonnet JF;
XX
DR   WPI; 2003-567944/53.
XX
PT   New immunogenic composition comprising a recombinant avipox virus that
PT   expresses in vivo in the animal the West Nile (WN) proteins prM, M or E,
PT   useful for inducing an immunological response against WN virus.
XX
PS   Example 10; Page 12; 24pp; English.
XX
CC   The invention relates to a novel immunogenic composition for inducing an
CC   immune response against West Nile fever virus (WNV) in an animal. The
CC   composition comprises a vehicle or excipient and a recombinant avipox
CC   virus that expresses in vivo in the animal the WNV proteins prM (pre-
CC   membrane protein), M (membrane protein) or E (envelope protein). The
CC   animal is selected from canine, feline, bovine, porcine, chicken, equine,
CC   a duck, a goose or a turkey. The composition of the invention
CC   demonstrates virucide activity and may be useful as a vaccine against
CC   WNV. The current sequence is that of the FC110 PCR primer of the
CC   invention which was used to amplify the plasmid pFC105
XX
SQ   Sequence 33 BP; 7 A; 7 C; 9 G; 10 T; 0 U; 0 Other;

  Query Match             82.6%;  Score 19;  DB 9;  Length 33;
  Best Local Similarity   100.0%;  Pred. No. 0.2;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATGACTGCAATTCCGGT 19
              |||||||||||||||||||
Db         28 TCATGACTGCAATTCCGGT 10


RESULT 14
ADM16860/c
ID   ADM16860 standard; DNA; 33 BP.
XX
```

```
AC   ADM16860;
XX
DT   20-MAY-2004  (first entry)
XX
DE   Plasmid pFC 105 PCR primer #1.
XX
KW   immunogen; vaccine; West Nile virus; ss; PCR; primer.
XX
OS   Synthetic.
XX
PN   US2004037848-A1.
XX
PD   26-FEB-2004.
XX
PF   26-FEB-2003; 2003US-00374953.
XX
PR   06-APR-2001; 2001US-0281923P.
PR   04-APR-2002; 2002US-00116298.
XX
PA   (AUDO/) AUDONNET J F.
PA   (MINK/) MINKE J M.
PA   (LOOS/) LOOSMORE S M.
PA   (KARA/) KARACA K.
XX
PI   Audonnet JF,  Minke JM,  Loosmore SM,  Karaca K;
XX
DR   WPI; 2004-191012/18.
XX
PT   Vaccine composition, useful in inducing an immune response against West
PT   Nile virus, comprises a vector that contains heterologous nucleic acid
PT   molecule(s), and that expresses in vivo in the animal a WNV protein.
XX
PS   Example 10; SEQ ID NO 18; 36pp; English.
XX
CC   The invention relates to an immunogenic or vaccine composition which
CC   induces an immune response against West Nile virus (WNV) in an animal
CC   susceptible to WNV comprises a vector that contains heterologous nucleic
CC   acid molecule(s) and that expresses in vivo in the animal a WNV E; WNV
CC   prM and E; WNV M and E; WNV prM, WNV M and E, WNV polyprotein prM-E, WNV
CC   polyprotein M-E, or WNV polyprotein prM-M-E. The composition is useful
CC   for inducing an immunological or protective immune response against WNV
CC   and against another pathogen of the animal. Also inducing an
CC   immunological or protective immune response against WNV in an animal
CC   comprises administering to the animal (a) the immunogenic or vaccine
CC   composition and (b) a WNV isolated antigen, immunogen or epitope, where
CC   (a) is administered prior to (b) in a prime-boost regimen, or (b) is
CC   administered prior to (a) in a prime-boost regimen, or (a) and (b) are
CC   administered together, either sequentially or in admixture. The present
CC   sequence is used in the exemplification of the invention.
XX
SQ   Sequence 33 BP; 7 A; 7 C; 9 G; 10 T; 0 U; 0 Other;

  Query Match             82.6%;  Score 19;  DB 12;  Length 33;
  Best Local Similarity   100.0%;  Pred. No. 0.2;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATGACTGCAATTCCGGT 19
              ||||||||||||||||||
Db         28 TCATGACTGCAATTCCGGT 10


RESULT 15
ABZ25431/c
ID   ABZ25431 standard; DNA; 36 BP.
XX
AC   ABZ25431;
XX
DT   27-MAR-2003  (first entry)
XX
DE   West Nile Virus PCR primer FC107, SEQ ID 9.
XX
KW   Virucide; vaccine; horse; dog; cat; cattle; pig; bird; West Nile virus;
KW   WNV; PCR; primer; ss.
XX
OS   West Nile Virus.
XX
PN   WO200281621-A2.
XX
PD   17-OCT-2002.
XX
PF   05-APR-2002; 2002WO-FR001200.
XX
PR   06-APR-2001; 2001FR-00004737.
XX
PA   (MERI-) MERIAL.
XX
PI   Loosmore SM,  Audonnet JF;
XX
DR   WPI; 2003-111799/10.
XX
PT   Vaccine for treatment or prevention of West Nile virus (WNV) infection,
PT   for use in veterinary medicine, comprises a recombinant virus expresssing
PT   a WNV structural protein.
XX
PS   Example 7; Page 31; 56pp; French.
XX
CC   The present invention relates to a vaccine for protecting horses, dogs,
CC   cats, cattle, pigs and birds against West Nile virus (WNV). The vaccine
CC   comprises: (i) one or more recombinant avipox, NYVAC or MVA viruses that
CC   express one of the WNV proteins prM, M and E and (ii) a vehicle or
CC   excipient. The present sequence is a PCR primer, which was used in an
CC   example from the invention
XX
SQ   Sequence 36 BP; 8 A; 6 C; 8 G; 14 T; 0 U; 0 Other;

  Query Match             82.6%;  Score 19;  DB 8;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 0.2;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATGACTGCAATTCCGGT 19
              ||||||||||||||||||
Db         31 TCATGACTGCAATTCCGGT 13


Search completed: September  6, 2005, 22:17:49
Job time : 234 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 6, 2005, 20:39:40 ; Search time 1492 Seconds
(without alignments)
746.964 Million cell updates/sec

Title: US-10-729-421-8
Perfect score: 23
Sequence: 1 tcatgactgcaattccggtcttt 23

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 4708233 seqs, 24227607955 residues

Word size : 10

Total number of hits satisfying chosen parameters: 417

Minimum DB seq length: 0
Maximum DB seq length: 60

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	20	87.0	36	6	AX589722	AX589722 Sequence
c	2	20	87.0	39	6	AX589721	AX589721 Sequence
c	3	19	82.6	33	6	AX589711	AX589711 Sequence
c	4	19	82.6	36	6	AX589702	AX589702 Sequence
c	5	19	82.6	56	6	AX224249	AX224249 Sequence
	6	12	52.2	25	6	AX182204	AX182204 Sequence
	7	12	52.2	25	6	AX382013	AX382013 Sequence
	8	12	52.2	34	6	BD173847	BD173847 JNK inhib
	9	12	52.2	51	6	AR444320	AR444320 Sequence
	10	12	52.2	51	6	AR444321	AR444321 Sequence
	11	12	52.2	60	6	CQ536043	CQ536043 Sequence
c	12	12	52.2	60	14	POLDIPI	M30219 Poliovirus
c	13	11	47.8	15	6	AR119503	AR119503 Sequence
c	14	11	47.8	16	6	AR285636	AR285636 Sequence
c	15	11	47.8	16	6	AR397627	AR397627 Sequence
	16	11	47.8	20	6	AR167035	AR167035 Sequence
	17	11	47.8	20	6	AR210690	AR210690 Sequence
c	18	11	47.8	20	6	AR301418	AR301418 Sequence
c	19	11	47.8	20	6	AR313575	AR313575 Sequence
	20	11	47.8	20	6	AX085449	AX085449 Sequence
	21	11	47.8	21	6	AR166997	AR166997 Sequence
	22	11	47.8	21	6	AR210652	AR210652 Sequence
	23	11	47.8	22	6	CQ768639	CQ768639 Sequence
	24	11	47.8	24	6	AX935032	AX935032 Sequence
c	25	11	47.8	25	6	AR364418	AR364418 Sequence
c	26	11	47.8	25	6	AR568240	AR568240 Sequence
c	27	11	47.8	26	6	AR542624	AR542624 Sequence
c	28	11	47.8	26	6	AX235895	AX235895 Sequence
	29	11	47.8	26	6	AX402749	AX402749 Sequence
c	30	11	47.8	27	6	BD183051	BD183051 Nucleic a
c	31	11	47.8	27	6	I22149	I22149 Sequence 8
c	32	11	47.8	28	6	I13959	I13959 Sequence 38
	33	11	47.8	30	6	AX085450	AX085450 Sequence
	34	11	47.8	30	6	AX101005	AX101005 Sequence
	35	11	47.8	33	6	AR004393	AR004393 Sequence
c	36	11	47.8	33	6	AR005205	AR005205 Sequence
	37	11	47.8	33	6	AR005206	AR005206 Sequence
	38	11	47.8	33	6	AR064955	AR064955 Sequence
c	39	11	47.8	33	6	AR072936	AR072936 Sequence
c	40	11	47.8	33	6	AR072938	AR072938 Sequence
	41	11	47.8	33	6	AR097185	AR097185 Sequence
	42	11	47.8	33	6	AR130683	AR130683 Sequence
	43	11	47.8	33	6	AR172032	AR172032 Sequence
	44	11	47.8	33	6	BD189149	BD189149 HCV Genom
	45	11	47.8	33	6	BD189296	BD189296 HCV Genom

ALIGNMENTS

RESULT 1
AX589722/c
LOCUS AX589722 36 bp DNA linear PAT 24-JAN-2003
DEFINITION Sequence 29 from Patent WO02081621.
ACCESSION AX589722
VERSION AX589722.1 GI:27901012
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
AUTHORS Loosmore,S.M. and Audonnet,J.C.
TITLE Vaccine against the nile fever virus
JOURNAL Patent: WO 02081621-A 29 17-OCT-2002;
MERIAL (FR)
FEATURES Location/Qualifiers
source 1..36
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="oligonucleotide"
ORIGIN

Query Match 87.0%; Score 20; DB 6; Length 36;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCATGACTGCAATTCCGGTC 20
||||||||||||||||||||
Db 31 TCATGACTGCAATTCCGGTC 12

RESULT 2
AX589721/c
LOCUS AX589721 39 bp DNA linear PAT 24-JAN-2003
DEFINITION Sequence 28 from Patent WO02081621.
ACCESSION AX589721
VERSION AX589721.1 GI:27901011
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.

```
REFERENCE   1
  AUTHORS   Loosmore,S.M. and Audonnet,J.C.
  TITLE     Vaccine against the nile fever virus
  JOURNAL   Patent: WO 02081621-A 28 17-OCT-2002;
            MERIAL (FR)
FEATURES             Location/Qualifiers
     source          1. .39
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="oligonucleotide"
ORIGIN

  Query Match             87.0%;  Score 20;  DB 6;  Length 39;
  Best Local Similarity   100.0%;  Pred. No. 0.19;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATGACTGCAATTCCGGTC 20
              ||||||||||||||||||||
Db         34 TCATGACTGCAATTCCGGTC 15


RESULT 3
AX589711/c
LOCUS       AX589711                33 bp    DNA     linear   PAT 24-JAN-2003
DEFINITION  Sequence 18 from Patent WO02081621.
ACCESSION   AX589711
VERSION     AX589711.1  GI:27901001
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Loosmore,S.M. and Audonnet,J.C.
  TITLE     Vaccine against the nile fever virus
  JOURNAL   Patent: WO 02081621-A 18 17-OCT-2002;
            MERIAL (FR)
FEATURES             Location/Qualifiers
     source          1. .33
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="oligonucleotide"
ORIGIN

  Query Match             82.6%;  Score 19;  DB 6;  Length 33;
  Best Local Similarity   100.0%;  Pred. No. 0.81;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATGACTGCAATTCCGGT 19
              |||||||||||||||||||
Db         28 TCATGACTGCAATTCCGGT 10


RESULT 4
AX589702/c
LOCUS       AX589702                36 bp    DNA     linear   PAT 24-JAN-2003
DEFINITION  Sequence 9 from Patent WO02081621.
ACCESSION   AX589702
VERSION     AX589702.1  GI:27900992
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Loosmore,S.M. and Audonnet,J.C.
  TITLE     Vaccine against the nile fever virus
  JOURNAL   Patent: WO 02081621-A 9 17-OCT-2002;
            MERIAL (FR)
FEATURES             Location/Qualifiers
     source          1. .36
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="oligonucleotide"
ORIGIN

  Query Match             82.6%;  Score 19;  DB 6;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 0.81;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATGACTGCAATTCCGGT 19
              |||||||||||||||||||
Db         31 TCATGACTGCAATTCCGGT 13


RESULT 5
AX224249/c
LOCUS       AX224249                56 bp    DNA     linear   PAT 10-SEP-2001
DEFINITION  Sequence 41 from Patent WO0160847.
ACCESSION   AX224249
VERSION     AX224249.1  GI:15554499
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Kinney,R.M., Kinney,C.Y., Butrapet,S., Gubler,D.L. and
            Bhamarapravati,N.
  TITLE     Avirulent, immunogenic flavivirus chimeras
  JOURNAL   Patent: WO 0160847-A 41 23-AUG-2001;
            The Secretary, Department of Health and Human Services (US)
FEATURES             Location/Qualifiers
     source          1. .56
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
ORIGIN

  Query Match             82.6%;  Score 19;  DB 6;  Length 56;
  Best Local Similarity   100.0%;  Pred. No. 0.79;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATGACTGCAATTCCGGT 19
              |||||||||||||||||||
Db         43 TCATGACTGCAATTCCGGT 25


RESULT 6
AX182204
LOCUS       AX182204                25 bp    DNA     linear   PAT 06-AUG-2001
DEFINITION  Sequence 14 from Patent WO0142441.
ACCESSION   AX182204
VERSION     AX182204.1  GI:15133479
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Reddy,S.I., Sadhu,L.I., Shukla,V.C. and Ferraiolo,G.I.
  TITLE     Plastid transformation
  JOURNAL   Patent: WO 0142441-A 14 14-JUN-2001;
            International Centre for Genetic Engineering and Biotechnology (IT)
FEATURES             Location/Qualifiers
     source          1. .25
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="PRIMER"
ORIGIN

  Query Match             52.2%;  Score 12;  DB 6;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 ATGACTGCAATT 14
```

```
                ||||||||||||
Db          1 ATGACTGCAATT 12


RESULT 7
AX382013
LOCUS       AX382013                  25 bp    DNA     linear   PAT 18-MAR-2002
DEFINITION  Sequence 17 from Patent WO0206497.
ACCESSION   AX382013
VERSION     AX382013.1  GI:19576835
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Reddy,V.S. and Sadhu,L.
  TITLE     Transplastomic plants
  JOURNAL   Patent: WO 0206497-A 17 24-JAN-2002;
            International Centre for Genetic Engineering and Biotechnology (IT)
FEATURES             Location/Qualifiers
     source          1. .25
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="PRIMER"
ORIGIN

  Query Match             52.2%;  Score 12;  DB 6;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 ATGACTGCAATT 14
                ||||||||||||
Db          1 ATGACTGCAATT 12


RESULT 8
BD173847
LOCUS       BD173847                  34 bp    DNA     linear   PAT 18-FEB-2003
DEFINITION  JNK inhibitor.
ACCESSION   BD173847
VERSION     BD173847.1  GI:28415180
KEYWORDS    WO 02062792-A/7.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 34)
  AUTHORS   Okawa,S., Naruo,K., Miwatashi,S., Kimura,H. and Kawamoto,T.
  TITLE     JNK inhibitor
  JOURNAL   Patent: WO 02062792-A 7 15-AUG-2002;
            TAKEDA CHEMICAL INDUSTRIES LTD,SHIGENORI OKAWA,KENICHI NARUO, SEIJI
            MIWATASHI,HIROYUKI KIMURA,TOMOHIRO KAWAMOTO
COMMENT     OS   Artificial Sequence
            PN   WO 02062792-A/7
            PD   15-AUG-2002
            PF   01-FEB-2002 WO 2002JP000828
            PR   02-FEB-2001 JP 01P   027570
            PI   SHIGENORI OKAWA,KENICHI NARUO,SEIJI MIWATASHI,HIROYUKI KIMURA,
            PI   TOMOHIRO KAWAMOTO
            PC   C07D417/04,C07D417/14,A61K31/4439,A61K31/4545,A61K31/506, PC
            A61P43/00,
            PC
            A61P1/18,A61P11/00,A61P17/00,A61P1/04,A61P7/00,A61P7/04,A61P21/ PC
             04,
            PC   A61P35/00,A61P35/02,A61P25/14,A61P9/10,A61P13/12,A61P27/06, PC
              A61P9/04,
            PC   A61P37/06,A61P29/00
            CC   PCR primer
            FH   Key             Location/Qualifiers
            FT   source          1. .34
            FT                   /organism='Artificial Sequence'.
FEATURES             Location/Qualifiers
     source          1. .34
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
ORIGIN

  Query Match             52.2%;  Score 12;  DB 6;  Length 34;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATGACTGCAA 12
                ||||||||||||
Db          9 TCATGACTGCAA 20


RESULT 9
AR444320
LOCUS       AR444320                  51 bp    DNA     linear   PAT 20-FEB-2004
DEFINITION  Sequence 731 from patent US 6670464.
ACCESSION   AR444320
VERSION     AR444320.1  GI:42672099
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 51)
  AUTHORS   Shimkets,R.A. and Leach,M.
  TITLE     Nucleic acids containing single nucleotide polymorphisms and
            methods of use thereof
  JOURNAL   Patent: US 6670464-A 731 30-DEC-2003;
FEATURES             Location/Qualifiers
     source          1. .51
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             52.2%;  Score 12;  DB 6;  Length 51;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 GACTGCAATTCC 16
                ||||||||||||
Db         12 GACTGCAATTCC 23


RESULT 10
AR444321
LOCUS       AR444321                  51 bp    DNA     linear   PAT 20-FEB-2004
DEFINITION  Sequence 732 from patent US 6670464.
ACCESSION   AR444321
VERSION     AR444321.1  GI:42672100
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 51)
  AUTHORS   Shimkets,R.A. and Leach,M.
  TITLE     Nucleic acids containing single nucleotide polymorphisms and
            methods of use thereof
  JOURNAL   Patent: US 6670464-A 732 30-DEC-2003;
FEATURES             Location/Qualifiers
     source          1. .51
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             52.2%;  Score 12;  DB 6;  Length 51;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 GACTGCAATTCC 16
                ||||||||||||
Db         12 GACTGCAATTCC 23
```

```
RESULT 11
CQ536043
LOCUS       CQ536043                  60 bp    DNA     linear   PAT 30-JAN-2004
DEFINITION  Sequence 5678 from Patent WO0210449.
ACCESSION   CQ536043
VERSION     CQ536043.1  GI:41502307
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Shoshan,A., Wasserman,A., Mintz,E., Mintz,L. and Faigler,S.
  TITLE     Oligonucleotide library for detecting rna transcripts and splice
            variants that populate a transcriptome
  JOURNAL   Patent: WO 0210449-A 5678 07-FEB-2002;
            Compugen Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .60
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match            52.2%;  Score 12;  DB 6;  Length 60;
  Best Local Similarity  100.0%;  Pred. No. 1.7e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 GACTGCAATTCC 16
              ||||||||||||
Db         30 GACTGCAATTCC 41


RESULT 12
POLDIPI/c
LOCUS       POLDIPI                   60 bp    DNA     linear   VRL 02-AUG-1993
DEFINITION  Poliovirus defective interfering particle 17 mRNA, partial cds.
ACCESSION   M30219
VERSION     M30219.1  GI:332915
KEYWORDS    .
SOURCE      Poliovirus
  ORGANISM  Poliovirus
            Viruses; ssRNA positive-strand viruses, no DNA stage;
            Picornaviridae; Enterovirus.
REFERENCE   1  (bases 1 to 60)
  AUTHORS   Kuge,S., Saito,I. and Nomoto,A.
  TITLE     Primary structure of poliovirus defective-interfering particle
            genomes and possible generation mechanisms of the particles
  JOURNAL   J. Mol. Biol. 192 (3), 473-487 (1986)
  MEDLINE   87169734
   PUBMED   3031313
COMMENT     Original source text: Poliovirus defective interfering particle 17,
            cDNA to viral RNA.
FEATURES             Location/Qualifiers
     source          1. .60
                     /organism="Poliovirus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:138953"
ORIGIN

  Query Match            52.2%;  Score 12;  DB 14;  Length 60;
  Best Local Similarity  100.0%;  Pred. No. 1.7e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 GACTGCAATTCC 16
              ||||||||||||
Db         33 GACTGCAATTCC 22


RESULT 13
AR119503/c
LOCUS       AR119503                  15 bp    DNA     linear   PAT 16-MAY-2001
DEFINITION  Sequence 26 from patent US 6153382.
ACCESSION   AR119503
VERSION     AR119503.1  GI:14102202
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 15)
  AUTHORS   Karn,J., Gait,M.John., Heaphy,S. and Dingwall,C.
  TITLE     Viral growth inhibition
  JOURNAL   Patent: US 6153382-A 26 28-NOV-2000;
FEATURES             Location/Qualifiers
     source          1. .15
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match            47.8%;  Score 11;  DB 6;  Length 15;
  Best Local Similarity  100.0%;  Pred. No. 7.8e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          11 AATTCCGGTCT 21
               ||||||||||
Db          14 AATTCCGGTCT 4


RESULT 14
AR285636/c
LOCUS       AR285636                  16 bp    RNA     linear   PAT 10-APR-2003
DEFINITION  Sequence 8 from patent US 6528640.
ACCESSION   AR285636
VERSION     AR285636.1  GI:29723230
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 16)
  AUTHORS   Beigelman,L., Burgin,A., Beaudry,A., Karpeisky,A.,
            Matulic-Adamic,J., Sweedler,D. and Zinnen,S.
  TITLE     Synthetic ribonucleic acids with RNAse activity
  JOURNAL   Patent: US 6528640-A 8 04-MAR-2003;
FEATURES             Location/Qualifiers
     source          1. .16
                     /organism="unknown"
                     /mol_type="unassigned RNA"
ORIGIN

  Query Match            47.8%;  Score 11;  DB 6;  Length 16;
  Best Local Similarity  100.0%;  Pred. No. 7.8e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy           9 GCAATTCCGGT 19
               ||||||||||
Db          13 GCAATTCCGGT 3


RESULT 15
AR397627/c
LOCUS       AR397627                  16 bp    RNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 8 from patent US 6617438.
ACCESSION   AR397627
VERSION     AR397627.1  GI:40134758
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 16)
  AUTHORS   Beigelman,L., Burgin,A.B., Beaudry,A., Karpeisky,A.,
            Matulic-Adamic,J., Sweedler,D. and Zinnen,S.
  TITLE     Oligoribonucleotides with enzymatic activity
  JOURNAL   Patent: US 6617438-A 8 09-SEP-2003;
FEATURES             Location/Qualifiers
     source          1. .16
```

```
                     /organism="unknown"
                     /mol_type="unassigned RNA"
ORIGIN

  Query Match           47.8%;  Score 11;  DB 6;  Length 16;
  Best Local Similarity 100.0%;  Pred. No. 7.8e+04;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          9 GCAATTCCGGT 19
              ||||||||||
Db         13 GCAATTCCGGT 3


Search completed: September  6, 2005, 22:42:49
Job time : 1494 secs
```